Trials & Tribulations

Studies on the fate, transparency and efficiency of clinical drug trials

DOCTORAL THESIS

C.A. (SANDER) VAN DEN BOGERT

Trials & Tribulations
Studies on the fate, transparency and efficiency of clinical drug trials

Proeven & beproevingen
Studies naar de bestemming, transparantie en efficiëntie van klinische geneesmiddelenstudies

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op woensdag 4 oktober 2017 des middags te 4.15 uur

door

Cornelis Andries van den Bogert
geboren op 22 oktober 1987
te 's-Hertogenbosch

TABLE OF CONTENTS

Chapter 1

Introduction

1

PROLOGUE

On March 13 2006, six healthy volunteers aged between 18 and 40 years were infused with a promising experimental drug, the so-called CD28 super-agonist Tegenero (TGN) 1412. It was the first time that the drug was tested in humans. Two other participants received sham (placebo) injections. This innovative mono-clonal antibody was developed to modulate the immune system in diseases due to chronic inflammation in various organ systems and hematological malignancies. Soon after the infusion, the six active drug recipients reported some worrying signs of headaches, rigors and myalgia. Other symptoms worsened quickly, including hypotension, tachycardia and fever. Three hours after the infusion, all alarms went off and the first high dose of corticosteroids was given. It was unfortunately only the beginning of the disease episode for these six young men, as it was followed by multi-organ failure and admission to an IC-unit for several weeks. They were eventually discharged with irreversible peripheral necrosis (charcoaled fingers) for one participant and a prolonged hematological and im-munological recovery for all six who received TGN1412.

The TGN1412 disaster, caused by overdose, is the nightmare of each investiga-tor conducting clinical trials on experimental, innovative drugs. Its aftermath produced several reports and recommendations, ranging from new European guidelines on conducting first in man (FIM) trials to the rather obvious recom-mendation that not all participants in a first in man trial should be treated with an experimental drug at the same time. Notwithstanding the reasonable point that the most cautious approach possible should be used when testing drug prototypes, the number of participants receiving TGN1412 in a 500-fold overdose is six or one does not alleviate the suffering for the exposed individual(s). It turned out that the health of these six participants was compromised because of some serious deficits of clinical research not limited to the TGN1412 trial. The TGN1412-disaster gave rise to some critical questions. Are clinical trials actually designed and conducted in a responsible way? Do they provide us with the right answers to the right questions? And are participants in trials sufficiently pro-tected for risks associated with testing unknown effects of new drugs? Clinical investigators, pharmaceutical companies, regulators employed by governments, journal editors, and trial participants might respond differently, but that these questions are still relevant 10 years after TGN1412 is very clear. Early 2016, a healthy volunteer did not survive a first in man trial with the compound BIA 10-2474, again due to overdose. This thesis aims to elucidate these challenges, in order to diagnose, and ultimately improve and sustain, the survival of clinical drug research.

SHORT OVERVIEW OF CLINICAL TRIAL EVOLUTION

Over the course of the twentieth century, clinical trials have become the gold standard for the evaluation of the effects of drugs in patients[1]. In modern society, doctors seek evidence from clinical trials to base their decisions on with regard to diagnosis and therapy. Furthermore, clinical trials are required for companies to obtain a license for their drug products. Medical practice has evolved to evidence based medicine, which places the randomized controlled trial (RCTs) and meta-analyses of RCTs on top of the hierarchy of evidence[2]. RCTs are considered as the most definitive, unbiased and preferable source to address clinical questions such as whether drugs should be used to treat patients with a given disease[3].

The emergence of the modern clinical trial[1] started in the Renaissance, around the 16th century. In this period, the groundwork of the modern (health) sciences was established, prioritizing experimental above anecdotal evidence. Several physicians, among which James Lind perhaps the most famous[2], started to conduct medical experiments to cure or prevent common diseases.

However, it would take the RCT until the 20th century to become common practice and the basis of health policy decision making. It proved its value when a research group (among which the statistician Austin Bradford Hill, 1897-1991) under the auspices of the British Medical Research Council successfully established the effectiveness of streptomycin and antibiotic supplementation against tuberculosis and tuberculosis resistance[4]. The findings of these trials were of direct benefit to clinical practice[5]. Meanwhile in the USA, Jonas Salk (1914-1995) successfully tested his killed virus polio vaccine, using a school for mentally incapacitated children in Pennsylvania as recruitment site for the first in man trial (informed consent was negotiated with state lawyers), and subsequently conducted a large-scale double-blind randomized field trial, involving 1.6 million children[6-8]. RCTs proved their value to demonstrate the therapeutic effect of drugs and governments adopted legislation[3] that benefit-risk profiles of new drugs should be substantiated by RCTs before companies could sell them on the market.

---

[1]   Comprehensively investigated by Dr. Laura Bothwell in *The emergence of the randomized controlled trial: origins to 1980. (Ph.D. thesis. New York: Columbia University, 2014.)*

[2]   Named after this famous Scottish navy doctor, the James Lind Library is a comprehensive online source of historical documents describing key developments of clinical trials (www.jameslindlibrary.org).

[3]   For example, the Kefauver-Harris Amendment (1962) in the US and the "Wet op de Genees-middelenvoorziening" (Drug Provision Act) in the Netherlands (enacted in 1958, coming into effect in 1963).

It became also clear that the conduct of clinical trials themselves also needed governance. Cases of unethical behavior, most notably the practices of the Nazi doctors in the concentration camps and the Tuskegee syphilis trial in the Southern United States pointed out that governance was needed to protect the people taking part in clinical trials[9]. Henry Beecher (1904-1976) described 22 more subtle examples of unethical clinical research, thereby underscoring the need for independent ethical review of studies involving humans[10]. The tension between science and ethics was recognized by the community, resulting in several ethical guidelines for clinical research (Nuremberg Code[11,12], Declaration of Helsinki[13], Belmont Report[14]), as well as legislation around the world requiring that investigators should follow these guidelines. The main message of these guidelines was that the informed consent, integrity, and (medical) interests of individual participants[4] involved in medical research should always trump the interests of the medical or scientific community. Institutional Review Boards (IRBs), or medical research ethics committees[5], were established, whose approval was required for all trials to be conducted in its hospital or research center. In the Netherlands, the first IRBs became operational around 1970, after the Dutch Health Council advised the Dutch government in 1955 on the importance of research ethics, study design, and adequate processing (integrity) of research data[15]. National requirements for clinical trials and IRBs were enshrined in the Dutch law in 1999. At the same time, a national authority for the accreditation of IRBs became operational: the Central Committee on Research Involving Human Subjects (Dutch abbreviation: CCMO). To harmonize the review process of drug trials in the increasingly globalized setting of science and drug development, guidelines were implemented on the level of the EU as well[16]. Furthermore, the International Conference on Harmonization (ICH) established the Good Clinical Practice (ICH-GCP) guideline, which aimed to set standards for clinical trials regarding ethical aspects and integrity of data for the three main markets the USA, EU, and Japan[17].

---

[4] Humans involved in clinical trials can be referred to as subjects, participants, volunteers, or in case they are included based on their disease, as patients. These definitions can be considered as more or less equivalent. For consistency purposes, participants will be used throughout this thesis. *Participant* highlights the role of the involved humans most appropriately with respect to the informed consent principle.

[5] Institutional Review Board (IRB) and Medical Research Ethics Committee (MREC) can both be used to describe expert groups that review the quality of the study protocol and whether the risks and burdens of a trial are proportionate to its benefits. IRB is the terminology used throughout the USA, whereas MREC is more common in some other parts of the world. IRB is the term used in this thesis.

THE FATE OF A TRIAL

Whether a clinical trial is an expensive, multimillion-dollar enterprise, involving centers, participants, regulators and investors around the world, or a small-scale academic initiative, all are designed with the intention to answer a clinical question of interest to the investigator. This question may be related to the efficacy of a new drug, toxicity, pharmacokinetics and -dynamics, biomarkers, or precision medicine. Furthermore, the rationale leading to this question can be purely driven by scientific or clinical curiosity, or can also be part of product development of a pharmaceutical company, to contribute to the licensing process. We can thus speak of the "scientific fate" of a trial, which concerns whether the trial has provided an answer to the research question. In addition to the scientific fate, trials conducted by the pharmaceutical industry also have a "(business) development" fate. This fate concerns whether the results of the trial support further development and marketing authorization of the drug. Failure to develop a marketable drug in the clinical phase of drug development means a heavy financial setback for the company.

The reason for distinguishing between the scientific and development fate of a trial is that the scientific fate should be independent of the direction (or magnitude) of the results, and that the development fate is determined by the direction of the results. If a trial demonstrates that a drug is not effective over placebo against the disease of interest, this may determine a disappointing development fate: the drug cannot be licensed for marketing. A disappointing development fate should not preclude the scientific fate of the trial from being completed and published in the literature. When evaluating the fate of drug trials, distinguishing between scientific failure, when the trial is discontinued or not reported, and development failure, when the trial shows no beneficial effect of the drug of interest, is of essence, as their causes and implications are different. Development failure in the clinical stage means loss of investment to the company and deferred hope for patients in need of new therapies. Although these consequences are highly unfortunate, such failures are part of the process. If a trial fails to provide an answer to the research question (or, fails to provide explanation why it did not find the answer), the investments in the trial are wasted and participants unnecessarily recruited and exposed to risks. Thus, the evaluation of the scientific fate of a trial should be, unlike the development fate, disconnected from the direction of results.

1
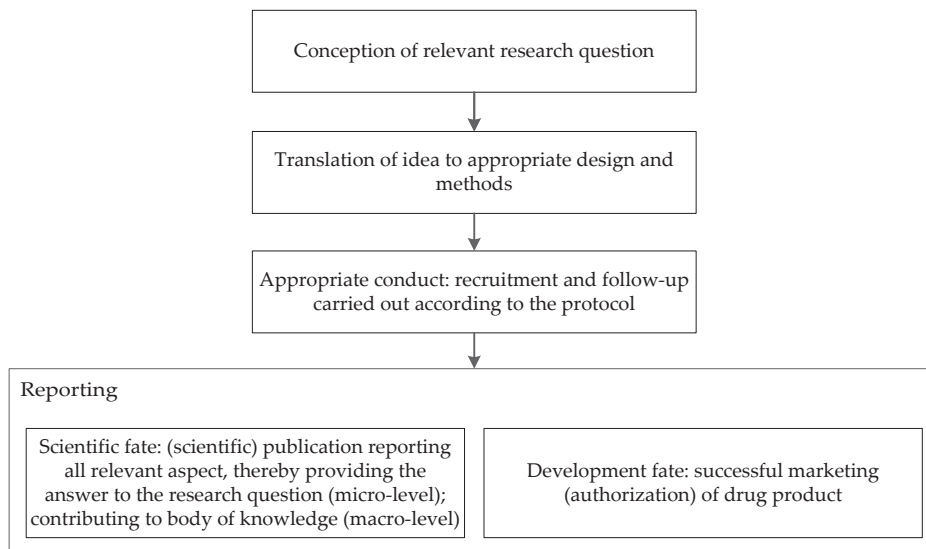
FIGURE 1.1 Framework for the optimal fate of clinical drug trials (adapted from Chalmers & Glasziou[18]).

Chalmers & Glasziou proposed a framework to identify research waste, which is closely connected to the fate of trials. Their framework distinguishes four stages: relevance of the research question, adequateness of the design and methods, accessibility of the publication and the usability of the report[18]. In addition to these four stages, the actual conduct of the trial is also relevant when judging the fate of a trial. Thus, adding the conduct as separate stage and merging the accessibility and usability of report to 'reporting', the conceptual framework for the fate of trials as illustrated in figure 1.1 will be used in this thesis.

A trial can only answer its research question if the right methods are chosen. In general, IRBs review the trial protocol, thereby ensuring that important choices regarding aspect such as control arms, randomization, blinding, use of placebo, and sample size are chosen in line with the research question[19]. Because of the IRB-review, we assume that the methodology of the trials is fit for purpose. However, for a particular type of trial, the first in man (FIM) trial, different paradigms exist regarding what should be the primary research question, or purpose. The classical view is that FIM trials should primarily aim to identify the maximum dose at which no toxicity occurs. More recently, others have proposed that the pharmacology (pharmacokinetics and –dynamics) should be the primary objective of FIM trials[20]. From these differences follows that there may be different views on the choice of methodology as well. This issue has received increased attention since 2006 (see the prologue), where it became clear that the classical view on FIM trials can increase the risk of safety events – by design. Ten years

later, it became clear that progress has been insufficient, when a healthy volunteer died after participating in a FIM trial on the new fatty acid amide hydrolase inhibitor *BIA 10-2474*. Regulatory authorities are currently increasingly adopting the pharmacologist view in their guidelines on FIM trials[21].

Next, trials can find an answer to their research questions if they are conducted and completed according to how they were designed. Completing a trial means that the needed number of participants was recruited, allocated and exposed to the test drugs, and gone through the planned follow-up procedures and measurements. Discontinuation of a trial before the planned end of recruitment and/or follow-up can imply that the research question cannot be answered[22], for example when insufficient participants were recruited or when the outcomes were measured with insufficient follow-up. Thus, in certain cases, discontinuation can be a suboptimal or even clearly unacceptable scientific fate of a clinical trial[23].

The appropriate scientific fate of all trials is a transparent publication in which all relevant aspects are reported and all results disclosed. If results of a trial are not reported, nobody beyond the selective group of investigators and sponsor will know the answer to the research question, and hence the fate of the trial is scientific death. Redundant or even dangerous subsequent trials may be conducted if investigators are unaware of previously conducted trials that have remained unpublished. It becomes more concerning if results are not (or selectively) being reported, depending on the direction or magnitude of the effect. This results in publication bias, which refers to the phenomenon that trials with positive results are more often published in the scientific literature than trials with negative results. The harmful implications of publication bias are not difficult to deduct. If the efficacy of drugs has been rejected by trials that have not been included in the meta-analyses, guidelines, textbooks or conference meetings on which physicians base the decision to treat (e.g. the recommendation to take a drug), this decision is inevitably misinformed and patients are unrightfully harmed by side effects. In 1645 the philosopher Francis Bacon already described the affection of scientists with positive results as an intrinsic human characteristic, and added his explicit disapproval of behaving according to this affection: "*Yet it is a proper and perpetual error in Humane Understanding, to be rather moved and stirred up by affirmatives than by negatives, although in truth it ought to be indifferent to both: Yet on the other hand the strength of a negative Instance is greater in constituting every Axiom*"[24]. In addition, fear of depreciation by sponsors and shareholders can motivate companies and investigators to silence negative results about their technology. A recent definition of publication bias is a "*tendency to submit (…) based on the direction or strength of findings*"[25]. Signs of publication bias in the medical literature were demonstrated by an investigation on drug applications to

the Finnish and Swedish drug regulatory authorities in 1980[26]. Some years later Dr. John Simes recognized the need for international trial registries to "*reviewing the clinical trial literature, which is free from publication bias*"[27]. However, publication bias turned out to be a persistent problem that has yet to be eradicated. A case study by the Swedish regulatory authority on RCTs investigating the new selective serotonin reuptake inhibitor antidepressants showed that most of the RCTs with negative findings have never been published in the scientific literature[28]. Since the terminology of publication bias was established in the eighties and nineties of the 20th century, people started to work on solutions in the early 2000's. Figure 1.2 shows the growing attention in literature towards publication bias from that time. The proposal of John Simes was finally operationalized with the launch of clinical trial registers where investigators could upload the design of their trial as well as the results after the trial was completed. In 2007, the Food and Drug Agency (FDA) made it mandatory for certain trials to use the trial platform clinicaltrials.gov for uploading results. The Dutch government updated legislation in 2011, which made the public sharing of trial results mandatory unless there were "motivated objections". Table 1.1 shows a list of global organizations involved in clinical trials that have issued statements aiming to achieve full reporting of all clinical trials in the public domain, demonstrating that the importance of the scientific fate of clinical trials is broadly recognized.

The different types of trials play a vital role in the development lifecycle of drug products, from (upstream) first in man trials until (downstream) post-marketing[29]. Thus, the development fate of a trial is whether it contributes to the lifecycle of a new or existing drug product. Before a drug can be marketed, a comprehensive clinical data package needs to be submitted to the drug marketing regulatory authorities (i.e. the regulatory product dossier), showing that the drug is efficacious against its target disease and has an acceptable toxicity profile. After the drug is licensed for marketing, the safety profile of the drug in the target population is further matured, through prospective post-marketing trials and observational studies. Efficacy in other patient populations can be investigated and, if successful, the marketing label can be extended. Furthermore, drug products are after the initial marketing authorization assessed by payers and health technology committees for their value in clinical practice, who also need clinical evidence to base their assessment on.
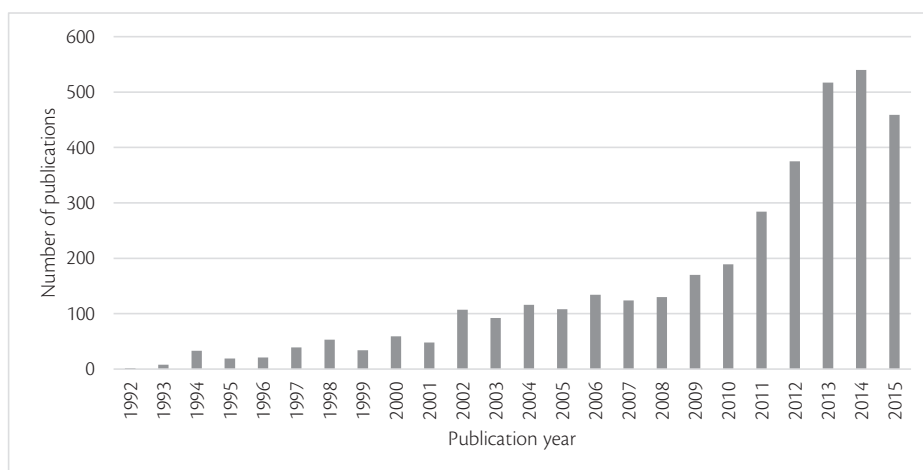
FIGURE 1.2 Publications indexed in PubMed under Mesh-term "Publication Bias" (date of search: 18 November 2016).

If the results of a trial do not favor the drug of interest, the further development and lifecycle of the drug may consequently be terminated. Companies may also disagree with the regulatory authorities on the interpretation of trial results, leading to a negative regulatory decision on the marketing authorization or reimbursement application. This can also mean termination of the lifecycle of the drug product, or at least substantial delay. This type of development success and failure of clinical drug development has been a topic of research in several previous theses. For example, one thesis investigated factors that were associated with success and failure in the initial marketing authorization application of new drug products[30]. Other theses investigated how the lifecycle of drug products continues once the market has been reached[31,32], or whether and how the safety profile of new (biological) drug products is further established in the post-marketing stage of the lifecycle[33-35]. These works have in common the interest in success and failure in the development lifecycle of drugs, and in the possibilities and limitations of the regulatory system and evidence-based medicine to influence this. What this thesis intends to add, is to prevent waste by focusing on the fate of drug trials, instead of drug products, drug side effects, diseases, or regulatory procedures. The main measurements of the development fate of trials will be inclusion in new marketing authorization application and the outcome of the application procedure.

TABLE 1.1 Selection of organizations and governments that have issued statements or legislation aiming at disclosure of clinical trial results. References to the source documents are provided in the first column.

Statement-issuing organization	Scope	Position on reporting
World Health Organization [36]	All interventional trials	Public reporting in general
International Committee of Medical Journal Editors [37]	All trials considered for publication in IMCJE-journal	Reporting design in registers
European Medicines Agency [38]	All clinical trials	Public reporting in general
Food and Drug Administration / National Institutes of Health / US Department of Health and Human Services [39,40]	"Applicable clinical trials" and all NIH-funded trials	Reporting results in registers
World Medical Association Declaration of Helsinki [13]	All clinical trials	Public reporting in general
International Federations representing the pharmaceutical industry [41]	Phase 3 trials	Scientific publication
European Organisation for Research and Treatment of Cancer [42]	All EORTC trials	Scientific publication and other venues

RISK-BASED GOVERNANCE

As outlined above, governance of clinical trials is in place to ensure that no unacceptable risks or burdens are imposed on the participants and that the data from the trials can be trusted. Regulations for governance originated in the aftermath of unethical research practices that affected vulnerable populations including children, ethnic minorities and women[9,36,37]. Governance was defined by Shaw et al as "the system of administration and supervision through which research is managed, participants and staff are protected, and accountability is assured. Governance is not the remit of any single institution (indeed, a guiding principle is that it is everyone's business)". Currently, ICH-GCP is the leading guideline for governance in most countries[17].

In addition to IRBs, governments installed inspectorate agencies to monitor compliance with ICH-GCP. These inspectorates (in the Netherlands, the Healthcare Inspectorate, abbreviated as IGZ) regularly visit clinical trial sites to check if the data are collected, processed and stored according to an ICH-GCP compliant plan. The decision regarding which trial site to visit can be made based on risk assessment. The various clinical trials within the jurisdiction of the inspectorate can be assessed and prioritized based on their risk profile, based on the idea that the risk of GCP issues is not randomly distributed, but differentially concentrated. Therefore, the distribution of resources and attention of the regulators should also not be random or equally distributed, but focused at the high end of the risk spectrum.

Investigators and IRBs may identify risks in trials in a systematic and structured way, and inspectorates may visit those trials that most urgently need checking of GCP compliance. Information about trials that is routinely collected may offer an opportunity to develop and validate a model able to identify such trials. Regulators and companies have developed such models for the purpose of trial oversight and supervision, but validation and publications of these models has been sparse. A complicating factor for relying on published models is that regulatory authorities need some level of discretion in their decision-making. There will always be circumstances by which a regulatory decision does not fit the standard model and hence a different choice is made. Nevertheless, routinely filled regulatory databases comprise a rich and promising source for a more structured approach to risk-based governance.

Not evaluated in this thesis is the effectiveness of regulatory processes or risk-based approaches themselves. Although the regulatory IRB database ToetsingOnline is used as tool, evaluating the clinical trial application and review process is not the intention. In ToetsingOnline, all clinical trials conducted in the Netherlands are registered from the moment of application to the IRB, as required per Dutch law. The research in this thesis identifies the fate of trials and indicators for risk-based supervision. In each chapter, suggestions and recommendations related to the research are provided, but the primary focus of this thesis will be diagnostic, not therapeutic. Theoretically, IRB databases are the best source for researching the fate of trials, as they contain all clinical trials within the jurisdiction of the IRB (not all types of trials are required to be registered other registries such as clinicaltrials.gov). However, since these databases are usually not developed with the intention of conducting research (which is the case for ToetsingOnline), their practical suitability is unknown. This will therefore be examined while investigating the fate of drug trials as outlined above, and recommendations are provided in the general discussion.

RESEARCH QUESTIONS AND THESIS OUTLINE

To summarize, the scientific and development fate of clinical drug trials is the primary interest of this thesis. The first research question is "what is the scientific and development fate of clinical drug trials, and what are the determinants of failure to reach the optimal fate?" In addition, this thesis aims to investigate possibilities and approaches to risk-based supervision of clinical trials. The second research question of this thesis is, therefore, "(how) can data that is routinely

registered in the clinical trial application and IRB-review process be used to develop a risk-based tool for clinical trial supervision?"

The major part of the research in this thesis is dedicated to the first research question, bundled in chapter 2. A cohort of all clinical drug trials reviewed by Dutch IRBs in 2007 was designed and followed from IRB review until their scientific fate (chapter 2.1-2.4), and until their development fate (chapter 2.5). The structure is chosen so that it follows the stages of progress of a drug trial: from design (chapter 2.1), conduct (chapter 2.2), to reporting and licensing (chapter 2.3, 2.4 and 2.5). As we first wrote and published the study protocol for the chapters 2.3 and 2.4, the initial idea, this protocol is included in Appendix 2.1. Chapter 3 quantitatively and qualitatively analyzes putative determinants of scientific death, i.e. why scientific studies are not (or selectively) reported, and proposes a theory of causal pathways. Furthermore, chapter 4 contains two analyses of drug development activity in the European context. Chapter 4.1 analyzes differences between phase 3 drug trials in oncology, a highly successful disease area in terms of pharmaceutical development versus psychiatry, a disease area in which the development fate of trials has been unsuccessful. Chapter 4.2 investigates whether the type of company and collaborating between companies are determinants of success in the marketing authorization application procedure. Chapter 5 investigates the literature and guidelines to identify indicators that can be used for risk-based supervision. In the general discussion (chapter 6), the answers to the research questions will be provided alongside with methodological reflections and recommendations for further research and practice. The discussion closes off with some personal reflections connected to the central research questions and an overall conclusion.

REFERENCES

1. Bothwell L. The emergence of the randomized controlled trial: origins to 1980. New York, USA: Columbia University; 2014.

2. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg.* Jul 2011;128(1):305-310.

3. Cochrane AL. *Effectiveness and efficiency: random reflections on health services.* London ,: Nuffield Provincial Hospitals Trust; 1972.

4. Daniels M, Hill AB. Chemotherapy of pulmonary tuberculosis in young adults; an analysis of the combined results of three Medical Research Council trials. *BMJ.* May 31 1952;1(4769):1162-1168.

5. Stewart SM, Crofton JW. The Clinical Significance of Low Degrees of Drug Resistance in Pulmonary Tuberculosis. *Am Rev Respir Dis.* Jun 1964;89:811-829.

6. Meldrum M. "A calculated risk": the Salk polio vaccine field trials of 1954. *BMJ.* Oct 31 1998;317(7167):1233-1236.

7. Salk JE. Studies in human subjects on active immunization against poliomyelitis. I. A preliminary report of experiments in progress. *J Am Med Assoc.* Mar 28 1953;151(13):1081-1098.

8. Guerrini A. Experimenting with humans and animals. *From Galen to animal rights.* Baltimore, Maryland, USA: The John Hopkins University Press; 2003.

9. Fairchild AL, Bayer R. Uses and abuses of Tuskegee. *Science.* May 07 1999;284(5416):919-921.

10. Beecher HK. Ethics and clinical research. *N Engl J Med.* Jun 16 1966;274(24):1354-1360.

11. Shuster E. Fifty years later: the significance of the Nuremberg Code. *N Engl J Med.* Nov 13 1997;337(20):1436-1440.

12. Bardy AH. Bias in reporting clinical trials. *Br J Clin Pharmacol.* Aug 1998;46(2):147-150.

13. van Lent M, Overbeke J, Out HJ. Recommendations for a uniform assessment of publication bias related to funding source. *BMC Med Res Methodol.* 2013;13:120.

14. The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, U.S. Department of Health Education and Welfare, Office of the Secretary. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. Washington DC, USA: U.S. Federal Government; 1979.

15. De Gezondheidsraad. Afschrift No.15/8 Betreffende proeven op mensen (in Dutch; aquired through the national archive of the Netherlands. The Hague: Gezondheidsraad; 1955.

16. The European Parliament, The Council of Europe. Directive 2001/20/EC of the European Parliament and of the Council on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use. Brussels, Belgium: Official Journal of the European Communities; 2001.

17. International Conference on Harmonisation. Tripartite Guideline. Guideline for Good Clinical Practice E6 (R1). Current Step 4 version. Geneva, Switzerland: ICH; 1996.

18. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet.* Jul 04 2009;374(9683):86-89.

19. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA.* Feb 01 1995;273(5):408-412.

1

20. Cohen A. Should we tolerate tolerability as an objective in early drug development? *Br J Clin Pharmacol.* Sep 2007;64(3):249-252.

21. European Medicines Agency. Concept paper on the revision of the Guideline on strategies to identify and mitigate risks for first-in-human clinical trials with investigational medicinal products (EMA/CHMP/446302/2016). London, UK: EMA; 2016.

22. Easterbrook PJ, Matthews DR. Fate of research studies. *J R Soc Med.* Feb 1992;85(2):71-76.

23. Briel M, Olu KK, von Elm E, et al. A systematic review of discontinued trials suggested that most reasons for recruitment failure were preventable. *J Clin Epidemiol.* Dec 2016;80:8-15.

24. Bacon F. Franc. Baconis de Verulamio / Summi Angliae Cancellarii /Novum organum scientiarum. [Francis Bacon of St. Albans Lord Chancellor of England. A 'New Instrument' for the sciences] Lugd. Bat: apud Adrianum Wiingaerde, et Franciscum Moiardum. Aphorism XLVI (pages 45-46). 1645.

25. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA.* Jan 15 1992;267(3):374-378.

26. Hemminki E. Study of information submitted by drug companies to licensing authorities. *BMJ.* Mar 22 1980;280(6217):833-836.

27. Simes RJ. Publication bias: the case for an international registry of clinical trials. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* Oct 1986;4(10):1529-1541.

28. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine--selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ.* May 31 2003;326(7400):1171-1173.

29. Sheiner LB. Learning versus confirming in clinical drug development. *Clin Pharmacol Ther.* Mar 1997;61(3):275-291.

30. Putzeist M. *Marketing authorisation of new medicines in the EU: towards evidence-based improvement*. Utrecht: Utrecht Institute for Pharmaceutical Sciences, Utrecht University; 2013.

31. Stolk P. *From new molecules to leads for innovation*. Utrecht: Utrecht Institute for Pharmaceutical Sciences, Utrecht University; 2008.

32. Langedijk J. *Continuous innovation in the drug life cycle*. Utrecht, the Netherlands: Utrecht Institute for Pharmaceutical Sciences, Utrecht University; 2016.

33. Vermeer NS. *Pharmacovigilance of biologicals: dynamics in post-approval safety learning*. Utrecht, the Netherlands: Utrecht Institute for Pharmaceutical Sciences, Utrecht University; 2015.

34. Duijnhoven RG. *Assessing long-term and rare adverse effects of medicines*. Utrecht, the Netherlands: Utrecht Institute of Pharmaceutical Sciences, Utrecht University; 2016.

35. Peeters PJHL. *Etiological studies in complex diseases*. Utrecht, the Netherlands: Utrecht Institute for Pharmaceutical Sciences, Utrecht University; 2016.

36. Dewar S, Boddington P. Returning to the Alder Hey report and its reporting: addressing confusions and improving inquiries. *J Med Ethics.* Oct 2004;30(5):463-469.

37. McNeill P. *The Ethics and Politics of Human Experimentation*. Cambridge: University Press; 1993.

Chapter 2

The fate of clinical drug trials

Chapter 2.1

Pharmacological versus classical approaches in the design of first in man clinical drug trials

Cornelis A. van den Bogert, Adam F. Cohen, Hubert G.M. Leufkens, Joop M.A. van Gerven

ABSTRACT

Objectives

To investigate the occurrence of pharmacokinetic (PK)/pharmacodynamic (PD) and tolerability approaches in first in man (FIM) trials in the Netherlands, and to evaluate whether this has changed in 2015 compared to 2007.

Methods

All FIM trials approved by all Dutch Institutional Review Boards (IRBs) in 2007 and in 2015 were selected. The original trial protocols, investigator's brochures and investigational medicinal product dossiers were the data sources. The design elements preclinical information, dose calculation, endpoints and dose escalation were assessed on the justification of the chosen approaches. The approach lacking PK/PD justification related to the preclinical information was only use of No observed adverse effect level (NOAEL) or No observed effect level (NOEL); related to dose calculation, unexplained allometric scaling; related to endpoints, not measuring PD parameters; and related to dose escalation, if escalation was only guided by safety/tolerability.

Results

In 2007, the Dutch IRBs approved 21 FIM trials and in 2015 they approved 34 FIM trials (55 in total). Seven out of 21 (33%) of the FIM trials from 2007, and 14 out of the 34 (41%) FIM trials from 2015 discussed only the NOAEL or NOEL as preclinical information. Furthermore, 5 of the 21 (24%) 2007 FIM trials and 12 of the 34 (35%) 2015 FIM trials used unexplained allometric scaling. PD parameters were measured in 15 of the 21 (71%) 2007 FIM trials and in 31 of the 34 (91%) of the 2015 FIM trials, and allometric scaling was only guided by safety/tolerability in 11 of the 20 (55%) dose escalation trials in 2007 and in 9 of the 33 (27%) dose escalation trials in 2015.

Conclusions

Trial protocols and investigator's brochures commonly lack PK/PD approaches for the design elements preclinical information, dose calculation, endpoints, and dose escalation. Contrary to preclinical information and dose calculation, PK/PD in endpoints and dose escalation seems to have increased. Regardless which approach is chosen, a structured justification of the design in FIM trial protocols is needed.

INTRODUCTION

Good clinical practice in clinical drug trials starts with choosing the optimal design. In first in man (FIM) trials, each first administration of a dosage is of interest because the response of the human body is never fully predictable. The endpoints for each dose, in the case of FIM trials pharmacokinetics (PK), pharmacodynamics (PD), or safety and tolerability, demonstrate whether investigation of additional or other doses are required to reach efficacy with acceptable hypothetical safety . Hence, dose and endpoints are two essential aspects when designing FIM trials regarding participant safety and successful clinical development[1-8].

Reports on the two major trial tragedies of the past decade acknowledged that the problems in these trials emerged from choices made in the study design, ignoring substantial risks[9-11]. Apparently, the guidelines by the Food and Drug Administration and the European Medicines Agency[4,12] did not provide sufficient guidance to prevent the poor design choices. As a response, approaches were proposed to rethink the classical view on phase 1 drug trials, and to reduce the risk of dose-related uncertainties by incorporating other dosing strategies and endpoints[1,5,13-19]. Proposed dose strategies included using the Minimal Anticipated Biological Effect Level (MABEL) in addition to the conventional No Observed Adverse Effect Level (NOAEL), consideration of interspecies pharmacokinetic and pharmacodynamic differences in addition to the allometric scales, and the abolition of irrational maximum tolerated dose (MTD)-testing. Furthermore, it was advocated that traditional safety/tolerability measurements needed to be augmented by pharmacodynamic endpoints as much as possible, as characterization of the relations between dose, exposure and effect may prevent dose-related harm.

Perhaps due to the scarce public availability of original trial protocols of FIM-studies, only a small amount of evidence exists on the use of "PK/PD-approaches" and "tolerability-approaches" in the design of FIM-trials. A pilot study on 7 trial protocols from 2009 suggested that this balance is still much on the side of the tolerability approach[15]. Our objective was, therefore, to investigate the occurrence of PK/PD and tolerability approaches in FIM trials in the Netherlands, and to evaluate the change in time.

METHODS

We selected all FIM trials reviewed in 2007 and in 2015 by the Dutch Institutional Review Boards (IRBs) from the database ToetsingOnline. This database ensured the inclusion of all trials, as submission of clinical trials through this portal is mandatory by law throughout the Netherlands. We excluded trials rejected by the IRB, trials that investigated generic products, biosimilars, new formulations of older drugs, and microdosing trials. We used the IRB-approved trial protocol, investigational medicinal product dossier, and investigator's brochure as the data sources for the analysis.

For each trial, we identified the design strategy for four elements: preclinical information, dose calculation, endpoints, and dose escalation. In the preclinical information part, we counted the frequency and proportion of trial protocols discussing NOAEL (tolerability approach), MABEL (PK/PD approach), and information from similar compounds. In the calculation element, we counted the frequency and proportion of trials that discussed only allometric scaling without PK/PD-based justification of the applied correction factor. We also counted frequencies and proportions of trials that discussed allometric scaling overall, additional PK-guided dose adjustments (for example, interspecies differences in metabolism), and/or additional PD-guided dose adjustments (for example, interspecies differences in target receptor affinity) in the calculation of the first dose. Regarding the endpoints, we separately counted the frequency and proportion of trials that included safety/tolerability, PK-parameters, and/or PD-parameters. Regarding dose escalation, we counted the frequency and proportion of dose escalation trials only guided by safety/tolerability, trials also guided by PK endpoints, and trials also by PD-endpoints.

The frequencies and proportions of the investigated design elements were graphically presented, stratified by the year of IRB-approval. Furthermore, because the FDA has published a separate guideline for oncology drug development which includes guidance on dose selection and escalation[20], we stratified the proportions also by oncology versus non-oncology. Many oncology drugs have a nonselective cytotoxic mechanism of action (for example, platinum agents, alkylating agents, antimetabolites, or taxanes[21]) which may justify non-pharmacologic approaches rather than drugs in other disease areas.

TABLE 2.1.1 Characteristics of the 55 first in man trials included in the analysis, stratified by year of IRB-review.

	Total (N=55)	2007 (N=21)	2015 (N=34)
Disease area			
Oncology	20 (36.4%)	3 (14.3%)	17 (50.0%)
Other disease area	35 (63.6%)	18 (85.7%)	17 (50.0%)
Drug type			
Small molecule	37 (67.3)	17 (81.0%)	20 (58.8%)
Peptide, antibody, ATMP	18 (32.7%)	4 (19.0%)	14 (41.2%)
Type participants			
Only healthy volunteers	29 (52.7%)	13 (61.9%)	16 (47.1%)
Only patients	22 (40.0%)	4 (19.0%)	18 (52.9%)
Mixed	4 (7.3%)	4 (19.0%)	0
Sample size			
Median number of planned participants (IQR)	66 (36-96)	64 (30-81)	66.5 (49.5-121.5)

Abbreviations: ATMP = advanced therapeutic medicinal product (cell, tissue or gene therapy); IQR = interquartile range.

RESULTS

In 2007, 21 FIM-trials were approved by the IRBs in the Netherlands, and 34 were approved in 2015. Table 2.1.1 summarizes the characteristics of the 55 FIM trials stratified by the year of approval. In 2015, the proportion of oncology FIM trials and of FIM trials investigating peptides, antibodies or advanced therapeutic medicinal products (ATMP) was larger compared to 2007. The increased proportion of FIM trials involving patients in 2015 can be explained by the increased proportion of oncology trials, in which it is more common practice to include patients than many other disease areas.

Figure 2.1.1 and 2.1.2 show the results for the four design strategy elements preclinical information, dose calculation, endpoints and dose escalation. The results sometimes differed numerically between 2007 and 2015 for some of the measurements, but the directions and magnitudes of the differences were inconsistent. The most substantial differences were found in the dose escalation (figure 2.1.2): 11 out of 20 (55%) trials in 2007 that performed dose escalation, escalated only based on safety/tolerability, compared to 9 out of 33 (27%) dose escalation trials in 2016. Furthermore, 8 out of 20 (40%) dose escalation trials were guided by PK-parameters in 2007 versus 24 out of 33 (73%) PK-guided dose escalation trials in 2015. Oncology FIM-trials seemed to use PD preclinical information and PD-based adjustment in the calculation of the dose more often compared to non-oncology trials. Twelve out of the 19 (63%) oncology trials used the MABEL approach from preclinical studies compared to 19 out of 36 (53%) of the non-oncology trials, and 11
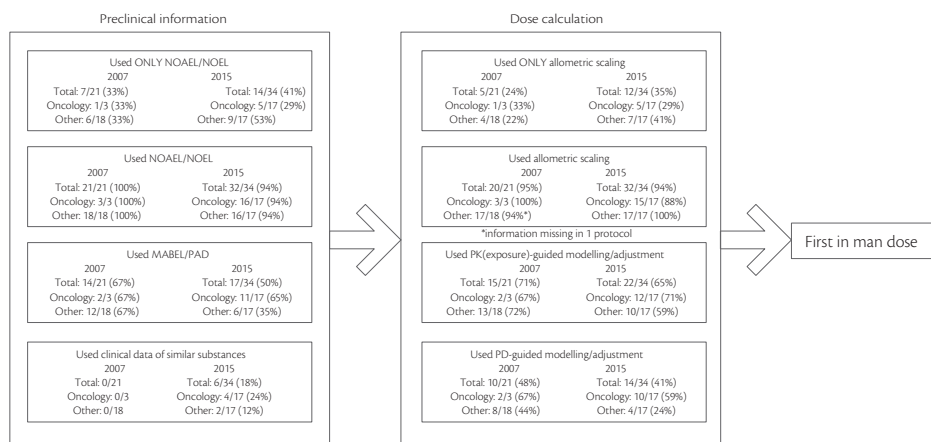
Preclinical information

Dose calculation

Used ONLY NOAEL/NOEL
2007	2015
Total: 7/21 (33%)	Total: 14/34 (41%)
Oncology: 1/3 (33%)	Oncology: 5/17 (29%)
Other: 6/18 (33%)	Other: 9/17 (53%)

Used NOAEL/NOEL
2007	2015
Total: 21/21 (100%)	Total: 32/34 (94%)
Oncology: 3/3 (100%)	Oncology: 16/17 (94%)
Other: 18/18 (100%)	Other: 16/17 (94%)

Used MABEL/PAD
2007	2015
Total: 14/21 (67%)	Total: 17/34 (50%)
Oncology: 2/3 (67%)	Oncology: 11/17 (65%)
Other: 12/18 (67%)	Other: 6/17 (35%)

Used clinical data of similar substances
2007	2015
Total: 0/21	Total: 6/34 (18%)
Oncology: 0/3	Oncology: 4/17 (24%)
Other: 0/18	Other: 2/17 (12%)

Used ONLY allometric scaling
2007	2015
Total: 5/21 (24%)	Total: 12/34 (35%)
Oncology: 1/3 (33%)	Oncology: 5/17 (29%)
Other: 4/18 (22%)	Other: 7/17 (41%)

Used allometric scaling
2007	2015
Total: 20/21 (95%)	Total: 32/34 (94%)
Oncology: 3/3 (100%)	Oncology: 15/17 (88%)
Other: 17/18 (94%*)	Other: 17/17 (100%)
*information missing in 1 protocol

Used PK(exposure)-guided modelling/adjustment
2007	2015
Total: 15/21 (71%)	Total: 22/34 (65%)
Oncology: 2/3 (67%)	Oncology: 12/17 (71%)
Other: 13/18 (72%)	Other: 10/17 (59%)

Used PD-guided modelling/adjustment
2007	2015
Total: 10/21 (48%)	Total: 14/34 (41%)
Oncology: 2/3 (67%)	Oncology: 10/17 (59%)
Other: 8/18 (44%)	Other: 4/17 (24%)

First in man dose

FIGURE 2.1.1 Approaches to use of preclinical information and to dose calculation. The figures are the numbers of trials that used the approach of that box divided by the number of trials in that stratum. Abbreviations: NOAEL: no observed adverse effect level; NOEL: no observed effect level (can be considered as similar to NOAEL); MABEL: minimal anticipated biological effect level; PAD: pharmacologically active level (can be considered as similar to MABEL); PK: pharmacokinetic; PD: pharmacodynamic.

out of 19 (58%) oncology trials used PD modelling or adjustment in dose calculation compared to 13 out of 36 (36%) non-oncology trials.

The preclinical information for the first dose determination was for almost all trials the NOAEL. Thirty-one out of 55 trials (56%) of the trials used the MABEL or Pharmacologically Active Dose (PAD) from preclinical studies. Furthermore, in 2015, 6 out of 34 FIM trials (18%) extrapolated clinical information from similar drugs.

Most of the trials (52 out of 55; 95%) used allometric scaling to calculate the first in human dose. In 17 out of 55 (31%) trials, no other methods than allometric scaling were described in the protocol to calculate the first in man dose. PK-guided modelling (for example, dose calculation based on predicted human oral absorption) was used in 37 out of 55 (67%) trials and PD-guided modelling (for example, dose calculation based on predicted human receptor occupancy) was used in 24 out of 55 (44%) trials.

Regarding the choice of endpoints, all trials measured safety/tolerability and PK parameters. Forty-six out of 55 (84%) trials also evaluated PD parameters. Two trials did not perform dose escalation; hence, we evaluated this in the 53 dose escalation trials. The decision to proceed to a next dose level was always guided by safety/tolerability parameters. In addition, PK parameters were taken into account in 32 out of 53 (60%), and PD parameters in 19 out of 53 (36%) dose escalation trials. In 2007, 20 (55%) and in 2015 9 (27%) dose escalation trials escalated only based on the safety/tolerability.
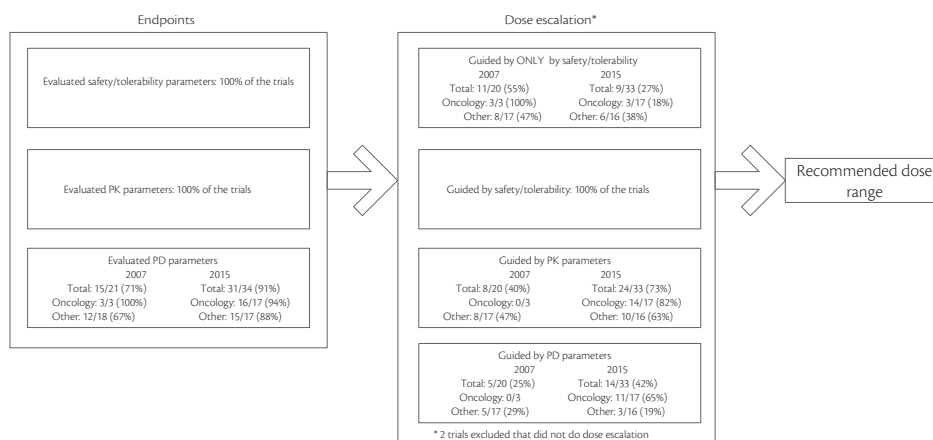
Endpoints

Dose escalation*

Evaluated safety/tolerability parameters: 100% of the trials

Evaluated PK parameters: 100% of the trials

Evaluated PD parameters
	2007	2015
Total:	15/21 (71%)	31/34 (91%)
Oncology:	3/3 (100%)	16/17 (94%)
Other:	12/18 (67%)	15/17 (88%)

Guided by ONLY by safety/tolerability
	2007	2015
Total:	11/20 (55%)	9/33 (27%)
Oncology:	3/3 (100%)	3/17 (18%)
Other:	8/17 (47%)	6/16 (38%)

Guided by safety/tolerability: 100% of the trials

Guided by PK parameters
	2007	2015
Total:	8/20 (40%)	24/33 (73%)
Oncology:	0/3	14/17 (82%)
Other:	8/17 (47%)	10/16 (63%)

Guided by PD parameters
	2007	2015
Total:	5/20 (25%)	14/33 (42%)
Oncology:	0/3	11/17 (65%)
Other:	5/17 (29%)	3/16 (19%)

Recommended dose range

* 2 trials excluded that did not do dose escalation

FIGURE 2.1.2 Endpoint and dose escalation approaches. The figures are the numbers of trials that used the approach of that box / the number of trials in that stratum. Abbreviations: PK = pharmacokinetic; PD = pharmacodynamic.

DISCUSSION

Our analysis of 55 FIM trial protocols found that the PK/PD approach in FIM trials seems to have increased in the guidance of dose escalation. Using the PK/PD-approach had not increased in the preclinical information, subsequent dose calculation, and choice of endpoints. Oncology trials seemed to use the PK/PD approach more often in preclinical information and dose calculation, and measured more often PD endpoints compared to non-oncology trials.

The presumed increase of the PK/PD-approach in dose escalation can be explained by the changing pharmaceutical pipeline which has become increasingly populated with targeted compounds[22]. PK/PD techniques and measurements such as labeling and biomarkers may be better available for these newer classes. We expected that the PK/PD approach in the use of preclinical information and dose calculation would have been increased as well. In the aftermath of TGN1412, regulators had published several guidelines[2,23] on its importance. It can be that PK/PD approaches have been compared to the conventional approach and made no difference, or that there was no PK/PD information available to use. However, in that case both approaches should at least have been mentioned and justified in the protocol and/or in the investigator's brochure. Regarding the dose escalation, it can be the case that the escalation was guided by PK/PD upfront, through limiting the dose escalation cohorts by pharmacological reasoning. We did not count these escalations as guided by PK/PD, as each dose increase in a FIM trial is in fact a new FIM trial for the new dose. The decision to progress to

a next dose should therefore always be transparently explained and justified by available PK and/or PD data.

A similar proportion of the oncology trials did use conventional approaches of only NOAEL, only allometric scaling, and only tolerability-guided escalation compared to non-oncology trials. However, the oncology trials did incorporate PD measurements (MABEL/PAD and PD endpoints) more often in the design compared to non-oncology trials. This can be explained by that oncology FIM trials include more often patients compared to non-oncology trials, making it possible to measure clinical endpoints. Furthermore, the wave of targeted therapy in oncology, is increasingly reaching the clinical development stage over the past ten years[24,25]. Contrary to the classical chemotherapies, biomarkers and disease biology play a central role in the discovery of these targeted therapies.

Among the preclinical sources of information, the NOAEL is the traditional, originating in the development of early cytostatic drugs in the 1960-1970's[26,27]. In the animal-human dose translation, the NOAEL is converted according to allometric scales to the Human Equivalent Dose (HED). The allometric scales have been developed on mathematical models predicting tissue exposure in humans based on animal data, adjusted for body surface area[28,29]. The HED is divided by a safety factor – by default 10, but may be increased or decreased based on case-by-case justification – to arrive at the maximum recommended starting dose (MRSD). As the unintended toxic effects of non-specific drugs (destroying functional DNA) were pharmacologically similar to the intended effects (destroying tumor DNA) and hence predictable, pre-clinical toxicity was an effective method to estimate the human pharmacological window. However, there are two problems with this approach. First, inter-species differences in absorption, target homology, target expression, and metabolism can make the extrapolation of animal toxicity completely irrelevant for humans[5,16,18]. The second problem will arise if the intended pharmacological mechanism of action of the drug is saturated at a much lower level than that toxicity occurs[15]. This toxicity is then most likely not caused by the intended pharmacological mechanism, and hence unpredictable with regard to location, timing, mechanism and severity. In these cases, toxicity is therefore not a suitable parameter to guide dose escalation. Preliminary data suggest that a more cautious pharmacology-based escalation approach should have been followed in the BIAL-102474-trial [9]. The intended pharmacological mechanism of action was saturated at a dose 20-fold lower compared to the dose in which the severe adverse events occurred. Escalation could have been stopped at a much lower dose, and the death of the participant might have been prevented.

To mitigate the safety risks related to dose-uncertainties in FIM-trials, we advocate that FIM trials use PK/PD approaches to justify the four design ele-

ments of preclinical information, dose calculation, endpoints and dose escalation. Whether and how these approaches are used should be determined on a case-by-case basis, and is determined by practical issues such as the availability of relevant biochemical parameters or tissue to measure PK/PD. A specific recommendation with regard to dose escalation is that the trial protocol should detail the PK and/or PD-based threshold until which the dose escalation is planned (for example, exposure below fraction X of the NOAEL, or below fraction Y of target receptor saturation). Some of the investigated trial protocols described this in detail, but other protocols provided no information on the escalation threshold. We were therefore unable to identify how the thresholds were established across the analyzed trials.

In all cases, investigators should provide (and IRBs should require) structured justification for approaches that are used as well as for approaches that are not used[30]. Efforts should be made to find and validate the best possible proxy measurements in case biochemical parameters are absent. In this way, the development trajectory from drug prototypes to drug treatments may become more efficient because drugs that do not induce the postulated PK/PD effects are identified in the earliest stage possible. The current guidelines support our arguments[3,4,12], but perhaps firmer regulatory oversight is needed to enforce further improvements.

In conclusion, PK/PD approaches to determine the first dose, endpoints and dose escalation in FIM trials are often not used, neither do trial protocols provide justification for not using them. The PK/PD approach seems to have become more common regarding dose escalation. The design choices of FIM trials differ on a case-by-case basis, but trial protocols should always provide a structured justification for (not) using the PK/PD approaches.

REFERENCES

1. Cohen AF, Burggraaf J, van Gerven JM, Moerland M, Groeneveld GJ. The use of biomarkers in human pharmacology (Phase I) studies. *Annu Rev Pharmacol Toxicol.* 2015;55:55-74.

2. European Medicines Agency. Guideline on Strategies to Identify and Mitigate Risks for First-in-Human Clinical Trials with Investigational Medicinal Products (EMEA/CHMP/SWP/28367/07) London, UK: EMA; 2007.

3. European Medicines Agency. Concept paper on the revision of the Guideline on strategies to identify and mitigate risks for first-in-human clinical trials with investigational medicinal products (EMA/CHMP/446302/2016). London, UK: EMA; 2016.

4. Food and Drug Administration. Guidance for Industry: Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers. Silver Spring, Maryland USA: U.S. Food and Drug Administration; 2005.

5. Cohen A. Pharmacokinetic and pharmacodynamic data to be derived from early-phase drug development: designing informative human pharmacology studies. *Clin Pharmacokinet.* 2008;47(6):373-381.

6. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov.* Aug 2004;3(8):711-715.

7. Lesko LJ, Atkinson AJ, Jr. Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. *Annu Rev Pharmacol Toxicol.* 2001;41:347-366.

8. Milligan PA, Brown MJ, Marchant B, et al. Model-based drug development: a rational approach to efficiently accelerate drug development. *Clin Pharmacol Ther.* Jun 2013;93(6):502-514.

9. The Temporary Specialist Scientific Committee "FAAH (Fatty Acid Amide Hydrolase)". Report on the causes of the accident during a Phase 1 clinical trial in Rennes in January 2016. . Saint Denis, France: Agence Nationale de Sécurité du Médicament et des produits de santé; 2016.

10. Expert Scientific Group mandated by the Secretary of State for Health. Phase One Clinical Trials Final Report. London, UK: The Stationery Office; 2006.

11. Randerson J. News & Comment: Fatal French clinical trial failed to check data before raising drug dose. *Nature* 2016.

12. Hume D. *A Treatise of Human Nature; Part 3: Of the will and direct passions.* London, UK1738.

13. Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. *Cochrane Database Syst Rev.* Feb 16 2017;2:MR000033.

14. Cohen A. Should we tolerate tolerability as an objective in early drug development? *Br J Clin Pharmacol.* Sep 2007;64(3):249-252.

15. Cohen AF. Developing drug prototypes: pharmacology replaces safety and tolerability? *Nat Rev Drug Discov.* Nov 2010;9(11):856-865.

16. Kenter MJ, Cohen AF. Establishing risk of human experimentation with drugs: lessons from TGN1412. *Lancet.* Oct 14 2006;368(9544):1387-1391.

17. Nair AB, Jacob S. A simple practice guide for dose conversion between animals and human. *J Basic Clin Pharm.* Mar 2016;7(2):27-31.

18.   Sharma V, McNeill JH. To scale or not to scale: the principles of dose extrapolation. *Br J Pharmacol*. Jul 2009;157(6):907-921.

19.   Swart P, Lozac'h F, Simon M, van Duijn E, Vaes WH. The impact of early human data on clinical development: there is time to win. *Drug Discov Today*. Jun 2016;21(6):873-879.

20.   Goldacre B. *Bad pharma : how drug companies mislead doctors and harm patients*. First American edition. ed. New York: Faber and Faber, Inc., an affiliate of Farrar, Straus and Giroux; 2013.

21.   Malhotra V, Perry MC. Classical chemotherapy: mechanisms, toxicities and the therapeutic window. *Cancer Biol Ther*. Jul-Aug 2003;2(4 Suppl 1):S2-4.

22.   Smietana K, Siatkowski M, Moller M. Trends in clinical success rates. *Nat Rev Drug Discov*. Jun 2016;15(6):379-380.

23.   Food and Drug Administration. Guidance for Industry: Immunogenicity Assessment for Therapeutic Protein Products. *Biotechnol Law Rep*. Jun 2013;32(3):172-185.

24.   Santos R, Ursu O, Gaulton A, et al. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov*. Jan 2017;16(1):19-34.

25.   Aggarwal S. Targeted cancer therapies. *Nat Rev Drug Discov*. Jun 2010;9(6):427-428.

26.   Freireich EJ, Gehan EA, Rall DP, Schmidt LH, Skipper HE. Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey, and man. *Cancer Chemother Rep*. May 1966;50(4):219-244.

27.   Schein PS, Davis RD, Carter S, Newman J, Schein DR, Rall DP. The evaluation of anticancer drugs in dogs and monkeys for the prediction of qualitative toxicities in man. *Clin Pharmacol Ther*. Jan-Feb 1970;11(1):3-40.

28.   Boxenbaum H. Interspecies scaling, allometry, physiological time, and the ground plan of pharmacokinetics. *J Pharmacokinet Biopharm*. Apr 1982;10(2):201-227.

29.   Zaharko DS, Dedrick RL, Bischoff KB, Longstreth JA, Oliverio VT. Methotrexate tissue distribution: prediction by a mathematical model. *J Natl Cancer Inst*. Apr 1971;46(4):775-784.

30.   Van den Bogert CA, Cohen AF. Need for a proactive and structured approach to risk analysis when designing phase I trials. *BMJ*. 2015;351:h3899.

2

Chapter 2.2

Recruitment failure and futility are the most common reasons for discontinuation of clinical drug trials

Cornelis A. van den Bogert; Patrick C. Souverein; Cecile T.M. Brekelmans; Susan W.J. Janssen; Gerard H. Koëter; Hubert G.M. Leufkens; Lex M. Bouter

ABSTRACT

Objectives

To identify the reasons for discontinuation of clinical drug trials and to evaluate whether efficacy-related discontinuations were adequately planned in the trial protocol.

Methods

All clinical drug trials in the Netherlands, reviewed by Institutional Review Boards (IRBs) in 2007, were followed until December 2015. Data were obtained through the database of the Dutch competent authority (CCMO) and a question-naire to the principal investigators. Reasons for trial discontinuation were the primary outcome of the study. Three reasons for discontinuation were analyzed separately: all-cause, recruitment failure, and efficacy-related (when an interim analysis had demonstrated futility or superiority). Among the efficacy-related discontinuations, we examined whether the data monitoring committee (DMC), stopping rule, and the moment of the interim analysis in the trial progress were specified in the trial protocol.

Results

Of the 574 trials, 102 (17.8%) were discontinued. The most common reasons were recruitment failure (33/574; 5.7%) and solely efficacy-related (30/574; 5.2%). Of the efficacy-related discontinuations, 10/30 (33.3%) of the trial protocols reported all three aspects in the trial protocol, and 20/30 (66.7%) reported at least one aspect in the trial protocol.

Conclusions

One out of five clinical drug trials is discontinued before the planned trial end, with recruitment failure and futility as the most common reasons. The target sample size of trials should be feasible, and interim analyses should be adequately described in trial protocols.

INTRODUCTION

Discontinuation of a clinical trial before completion of the planned recruitment and data collection can be the best decision for the trial participants. This is clearly the case if unexpected severe adverse events emerge in one or more trial arms. For example, the Cardiac Arrhythmia Suppression Trial (CAST) trial was discontinued after an interim analysis showed a higher mortality rate in the active drug arms compared to the placebo arm [1]. Similarly, a planned interim analysis of the primary outcome of a trial can conclusively demonstrate the futility or superiority of one of the trial arms before the end of follow-up. The ethical principle of equipoise is then violated and the trial should be discontinued [2,3]. However, concerns exist about whether these interim analyses are in practice adequately planned, conducted and interpreted [4-6].

Discontinuation for commercial reasons can be at odds with sound methodology, as for example when an interim analysis was not planned or not performed according to the trial protocol. The likelihood is then increased that a chance finding in the interim analysis leads to a wrong decision to discontinue [7]. The International Conference on Harmonization established guidelines on these issues [8], specifying that clear stopping rules and the moment in the trial progress (at a specified number of included participants or number of events) should be defined, and that a Data Monitoring Committee (DMC) should be in place to perform the interim analysis. The European Clinical Trial Regulation (coming into effect as of 2018) also clearly states the importance of describing possible interim analyses in full detail in the trial protocol [9].

The occurrence and determinants of discontinuation of clinical trials has been empirically investigated in various settings [10,11], but this research may need to be updated as the samples were small and/or their findings may be outdated. Therefore, we investigated the frequency and reasons for discontinuation of clinical drug trials among an inception cohort of clinical drug trials and identified determinants for the most common reasons for discontinuation. Furthermore, we evaluated whether discontinuations after an interim analysis demonstrating either futility or superiority did so according to the trial protocol.

METHODS

The current study is a follow-up analysis of an inception cohort of all clinical drug trials reviewed by one of the accredited Institutional Review Boards (IRB) in the Netherlands in 2007. The design of this study has been published before [12], as

well as the results of which trials in the cohort were published in the scientific literature [13]. The data source was ToetsingOnline, the database maintained by the Central Committee on Research Involving Human Subjects (Dutch abbreviation: CCMO) that contains all IRB-reviewed clinical trials in the Netherlands. Other data sources were the complete trial files that were submitted to the CCMO in its role as national competent authority [14], including the original trial protocols submitted to the IRBs, the end-of-trial forms that investigators must submit when the study has ended (the EudraCT B7-form).

All drug trials (both randomized and non-randomized), reviewed by a Dutch IRB in 2007 (n= 622, figure 2.2.1) were identified and followed until December 2015 (the end of the study period). Trials that were rejected by the IRB (n=19), never started recruitment (n=19), or were still running at the time of data collection (n=10) were excluded from the analysis. Hence, 574 trials were selected for this study.
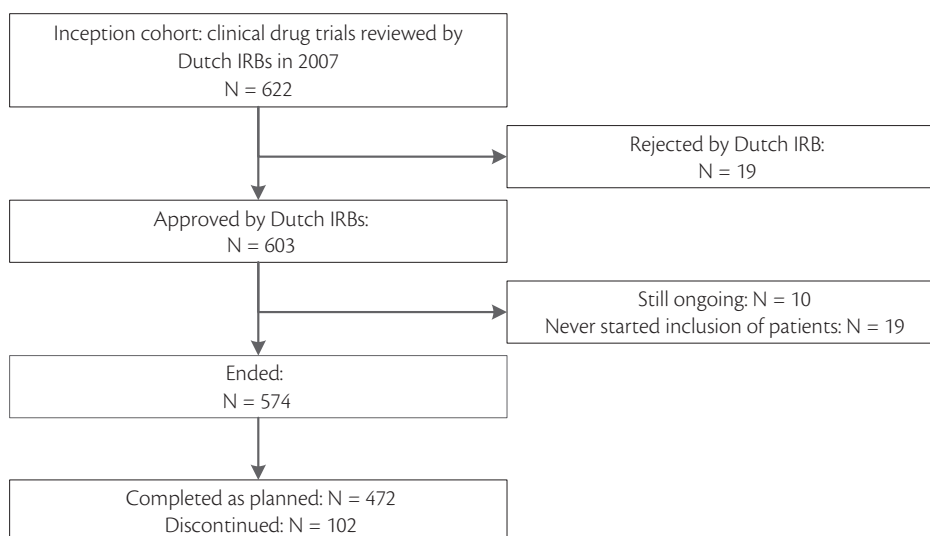
FIGURE 2.2.1. Selection of analyzed trials from inception cohort

We used investigator-reported information about the end of trial to the IRB and to the CCMO to classify whether they were discontinued or completed as planned, and to classify the reason for discontinuation. The first source was the EudraCT End-of-Trial form (also coded as the B7-form). This form, which is used by clinical trial authorities throughout the EU, requires investigators to report whether the trial was completed as planned or discontinued. In case of discontinuation, investigators must provide on this form one or more pre-specified reasons for discontinuation (the first version), or write other reasons in an open text box (the

second version). If this form was missing or incomplete in the CCMO archive, we searched for other sources in the clinical trial dossier, such as e-mail correspondence between investigators and the IRB, notifying the end of trial. We also used information from a questionnaire sent to all principal investigators. Questionnaires (see appendix 2.2 for more information and the questionnaire templates) were e-mailed to the principal investigators (PIs) of the trials, asking for reasons for non-publication for another analysis of the cohort[12], and whether the trial was completed as planned or discontinued, if the other sources were unavailable. If the PI had left the company or the hospital that conducted the trial, we tried to contact the PI at his current affiliation, or otherwise we attempted to contact colleagues of the PI that were involved in the same trial. After location of the right person, at maximum two reminders were sent. All Dutch accredited IRBs were asked for permission to send the questionnaire to the PIs. All IRBs consented and provided a signed letter of endorsement, which we attached to the questionnaire. The list of 23 Dutch accredited IRBs can be found on the website of the CCMO[15]. The end-of-trial form was missing of 186 of the 574 (32%) trials that were included in the analysis. Principal investigators of 73 of these trials responded to our questionnaire, completing the information on the end-of-trial. Of the remaining 113, of 87 trials we found other documents than the end-of-trial form indicating that the trial had started (for example, emails from the IRB or amendments), or we found that the trial was published. Of 26 trials, the IRB dossier did not contain information about the completion status and were nonresponding to the questionnaire [13]. After review of these 26 trials by two authors (CAB and CTMB), we decided that it would be most reasonable to consider these 26 trials as being completed as planned. In the Netherlands, it is common practice only to report to the IRB in case of irregularities such as discontinuation. Thus, we decided that it would be most reasonable to assume that all discontinuations had been reported to the IRB and/or by the questionnaire, and that trials with missing end-of-trial information were completed as planned. Reasons for discontinuation and their classification (in case they were reported in open-text format) were collected in a data extraction document in duplo by one investigator (CAB), double-checked by a research assistant. Differences were solved by consensus.

The investigator-reported reason(s) for discontinuation was the main outcome of the study. We categorized the reasons according to the pre-specified categories on the B7-form. Reasons reported in the open text box that could not be reclassified into the pre-specified reasons were described separately. Trials could be counted several times if investigators reported more than one reason for discontinuation.

Candidate determinants were trial characteristics planned target sample size, sponsor, phase, centers involved, randomization, and the disease area. These characteristics are filled out by investigators on a standard form for the IRB trial application, which is mandatory and identical throughout the country.

First, the frequencies of all reported reasons for discontinuation were described. Three dichotomous discontinuation outcomes were defined for further analysis: all-cause discontinuation, discontinuation due to recruitment failure, and discontinuation because an interim analysis demonstrated futility or superiority (efficacy related). All discontinuations reporting recruitment failure among the reasons were classified as such, because we judged reasons reported together with recruitment failure to be related to the recruitment failure. Discontinuations were only classified as efficacy-related if no other reasons (e.g. safety issues) were reported. This was done because the goal was to analyze determinants for trials solely discontinued because of an interim analysis that demonstrated futility or superiority. If other reasons, such as safety issues, were reported, the role of the interim analysis for futility or superiority may have been trivial compared to the other reasons for the decision to discontinue the trial. Percentages were described for all trial characteristic categories of these three discontinuation outcomes (all-cause, recruitment failure, and efficacy-related), and for trials that were completed as planned.

Furthermore, among the efficacy-related discontinuations we examined the trial protocol if the interim analysis was planned. We examined three aspects that should be described in the trial protocol according to the ICH-guideline [8]: mentioning a DMC, specification of the stopping rule, and specification of the moment (number of included participants and/or number of primary outcome events) of the interim analysis. We calculated the proportion of trials discontinued for efficacy covering at least one of these aspects in their trial protocol.

We used multivariable Poisson regression analysis to evaluate the association of trial characteristics with all-cause, efficacy-related and inclusion failure-related discontinuation. The crude and adjusted incidence rate ratio (IRR) and 95% confidence interval (CI) were estimated in three models; one with the outcome all-cause discontinuation, one with the outcome efficacy-related discontinuation, and one with the outcome discontinuation due to recruitment failure. All trials were included in the all-cause discontinuation model, and the trial characteristics sample size, sponsor, phase, centers, randomization, and disease area were tested. Only the phase 2 and phase 3 trials were included in the efficacy-related discontinuation model, as phase 1, phase 4 and other than phase 1-4 trials often do not measure efficacy and are therefore in general not at risk for efficacy-related discontinuation. In the efficacy-related discontinua-

tion model, the characteristics sample size and disease area were tested, based on the descriptive numbers. Phase 1 trials were excluded from the recruitment failure model, because these trials have different recruitment strategies (often healthy volunteers), face different recruitment challenges, and should therefore not be included in the multivariable model. In the recruitment failure model, we tested the characteristics sample size and sponsor, to look if we could replicate the findings of a previous study [10]. For the multivariable analysis, we merged the following trial characteristic categories to one category: investigator-initiated trials with and without industry (co-)funding (to investigator-initiated trials); national and international multicenter trials (to multicenter), the trial phases 2, 3, 4, and other than phase 1-4 (to other than phase 1), and the disease areas other than oncology (to other than oncology; as oncology trials include patients who are typically very ill and are therefore interesting to compare against the other disease areas). The multivariable analysis was done in Stata version 14.1.

RESULTS

Of the 574 analyzed trials, 472 were completed as planned and 102 (18%) were discontinued by December 2015 (figure 2.2.1). Table 2.2.1 summarizes the characteristics of the included trials, and table 2.2.2 describes the reasons for discontinuation as reported by the investigators. The most frequent reason was recruitment failure (no or slow recruitment): of the 102 discontinued trials, 33 (32%) were discontinued for this reason (or 5.7% of the total number of 574 trials), followed by 31 trials (30%) that were discontinued for futility as demonstrated by an interim analysis (5.4% of the total).

Thirty discontinuations (5.2%) were solely efficacy-related and thus should have been based on a planned interim analysis. Twenty trials (67% of the solely efficacy-related discontinuations) were discontinued while not describing all three essential aspects of an interim analysis (a DMC, the moment of the interim analysis in the trial and the stopping rules) in the protocol. Planning of the stopping rules was the aspect that was most often missing (in 18 (40%) of these protocols).

Table 2.2.3 shows the percentages of the trial characteristics for all-cause, solely efficacy-related and recruitment failure discontinuations. The results of the multivariable analysis are shown in the supplementary results: table S2.2.1-S2.2.3. Almost all trials that were discontinued solely efficacy-related were industry-sponsored (29 industry-sponsored and one investigator-initiated, table 2.2.3). Because there was only one efficacy-related discontinuation among investigator-initiated trials, the sponsorship variable was not included in the

TABLE 2.2.1 Characteristics of the trials included in the analysis.

Sample size		
Planned target sample size, median (IQR)	72 (25-320)	
Sponsor		
Pharmaceutical industry	352	61.3%
Investigator (industry (co-)funded)	71	12.4%
Investigator (no industry funding involved)	151	26.3%
Phase		
Phase 1	119	20.7%
Phase 2	130	22.6%
Phase 3	172	30.0%
Phase 4	57	9.9%
Other than phase 1-4*	96	16.7%
Centers		
Single center	249	43.4%
Multi center only in the Netherlands	54	9.4%
Multi center in the Netherlands and the EU	82	14.3%
Multi center in the Netherlands and outside the EU	189	32.9%
Randomization		
Randomized trial	418	72.8%
Non-randomized trial	156	27.2%
Disease area		
Oncology	113	19.7%
Neurological and psychiatric diseases	109	19.0%
Cardiovascular diseases	62	10.8%
Endocrine diseases	58	10.1%
Infectious diseases	42	7.3%
Other	190	33.1%

* Trials carried out using medicinal products in connection with objectives other than those referred to in the phase definitions 1-4. Such trials are not intended primarily to provide information about the product itself, but a medicinal product is needed in order to address the objective of the trial.

multivariable model for solely efficacy-related discontinuation (table S2.2.1). Investigator-initiated trials were associated with discontinuation due to recruitment failure: 23 (10.4%) of the 222 investigator-initiated vs. 10 (2.8%) of the 352 industry-sponsored trials were discontinued due to recruitment failure (adjusted IRR 2.0; 95% CI 0.9-4.6, table S2.2.3). The association was not statistically significant in the multivariable analysis due to the low numbers.

TABLE 2.2.2 Frequencies and percentages of the reported reasons for discontinuation.

	Frequency reason was reported*	% of the discontinued trials (N = 102)	% of the full sample (N = 574)
After interim analysis that should have been planned			
Interim analysis demonstrated futility	31	30.4	5.4
Interim analysis demonstrated superiority	2	2.0	0.3
Solely efficacy-related[†]	30	29.4	5.2
Trial protocol specified DMC[†]	15	14.7	2.6
Trial protocol specified stopping rules[‡]	12	11.8	2.1
Trial protocol specified the moment of the interim analysis in the trial progress[‡]	18	17.6	3.1
Trial protocol specified all 3 aspects[‡]	10	33.3	1.7
Trial protocol specified at least one of the 3 aspects[‡]	20	19.6	3.5
After interim analysis that could not have been planned			
Interim analysis due to safety signals	14	13.7	2.4
Interim analysis because results from other trials became available	2	2.0	0.3
Other reasons			
Recruitment failure	33	32.4	5.7
Financial issues	10	9.8	1.7
Product manufacturing or regulatory issues	4	3.9	0.7
Only Dutch sites closed, international trial continued	2	2.0	0.3
Unfeasible pharmacokinetics	1	1.0	0.2
Suspension of trial after GCP-inspection	1	1.0	0.2
Organizational issues	1	1.0	0.2
Reason missing	5	4.9	0.8

GCP = Good Clinical Practice

* 93 trials reported one reason, 4 trials reported 2 different reasons, and 5 trials only reported discontinuation, but not the reason

[†] Only examined among the protocols of the 30 trials that were discontinued solely efficacy-related.

[‡] Solely efficacy-related was after interim analysis demonstrated either futility or superiority. Three trials were excluded because reporting also other reasons than interim analysis demonstrating futility or superiority. Two of these three trials reported discontinuation after an interim analysis due to safety signals, and one trial reported recruitment failure as other reasons for discontinuation.

TABLE 2.2.3 Proportion of clinical drug trials discontinued (all-cause, solely efficacy-related, and recruitment failure), stratified by trial characteristics.

	Completed as planned	Discontinued (all-cause)	Discontinued for efficacy[†]	Discontinued for recruitment
All trials (n = 574)	N = 472 (82.2%)	N = 102 (17.8%)	N = 30 (5.2%)	N = 33 (5.7%)
Sample size				
Planned target sample size, median (IQR)	68 (24-314)	120 (40-392)	309 (78-635)	78 (23-180)
Sponsor				
Pharmaceutical industry (N = 352)	288 (81.8%)	64 (18.2%)	29 (8.2%)	10 (2.8%)
Investigator (industry (co-)funded) (N= 71)	56 (78.9%)	15 (21.1%)	0 (0%)	8 (11.3%)
Investigator (no industry funding involved) (N = 151)	128 (84.8%)	23 (15.2%)	1 (0.7%)	15 (9.9%)
Phase				
Phase 1 (N = 119)	108 (90.8%)	11 (9.2%)	1 (0.8%)	2 (1.7%)
Phase 2 (N = 130)	98 (75.4%)	32 (24.6%)	16 (12.3%)	9 (6.9%)
Phase 3 (N = 172)	133 (77.3%)	39 (22.7%)	13 (7.6%)	12 (7.0%)
Phase 4 (N = 57)	45 (78.9%)	12 (21.1%)	0 (0%)	7 (12.3%)
Other than phase 1-4* (N = 96)	88 (91.7%)	8 8.3%)	0 (0%)	3 (3.1%)
Centers				
Single center (N = 249)	219 (88.0%)	30 (12.0%)	1 (0.4%)	18 (7.2%)
Multi center only in the Netherlands (N = 54)	43 (79.6%)	11 (20.4%)	1 (1.9%)	7 (13.0%)
Multi center in the Netherlands and the EU (N = 82)	68 (82.9%)	14 (17.1%)	3 (3.7%)	2 (2.4%)
Multi center in the Netherlands and outside the EU (N = 189)	142 (75.1%)	47 (24.9%)	25 (13.2%)	6 (3.2%)
Randomization				
Randomized trial (N = 418)	344 (82.3%)	74 (17.7%)	23 (5.5%)	22 (5.3%)
Non-randomized trial (N = 156)	128 (82.1%)	28 (17.9%)	7 (4.5%)	11 (7.1%)
Disease area				
Oncology (N = 113)	81 (71.7%)	32 (28.3%)	15 (13.3%)	7 (6.2%)
Neurological and psychiatric diseases (N = 109)	93 (85.3%)	16 (14.7%)	3 (2.8%)	5 (4.6%)
Cardiovascular diseases (N = 62)	52 (83.9%)	10 (16.1%)	3 (4.8%)	3 (4.8%)
Endocrine diseases (N = 58)	47 (81.0%)	11 (19.0%)	3 (5.2%)	1 (1.7%)
Infectious diseases (N = 42)	38 (90.5%)	4 (9.5%)	0 (0%)	1 (2.4%)
Other (N = 190)	161 (84.7%)	29 (15.3%)	6 (3.2%)	16 (8.4%)

IQR = interquartile range

* Trials carried out using medicinal products in connection with objectives other than those referred to in the phase definitions 1-4. Such trials are not intended primarily to provide information about the product itself, but a medicinal product is needed in order to address the objective of the trial.

[†] Solely efficacy-related was after interim analysis demonstrated either futility or superiority. Three trials reporting futility were not defined as solely efficacy-related because they reported also other reasons than interim analysis demonstrating futility or superiority. Two of these three trials reported discontinuation after an interim analysis due to safety signals, and one trial reported recruitment failure as other reasons for discontinuation.

Another determinant for both efficacy-related discontinuation and discontinuation due to recruitment failure is the number and location of centers involved. Multicenter trials also conducted outside the EU had a significantly higher likelihood of efficacy-related discontinuation compared to single- and multicenter trials within the Netherlands or the EU (13% vs. 0.4-4%, table 2.2.3), whereas single center trials and multicenter trials only in the Netherlands had a higher likelihood of discontinuation due to recruitment failure compared to multicenter trials outside the Netherlands (7-13% vs. 2-3%, table 2.2.3). These findings could be explained by the fact that most of the international multicenter trials were industry-sponsored phase 3 trials, and that most of the non-phase 1 single center trials were investigator-initiated. Because of this multicollinearity with sponsorship and trial phase, we did not include the center variable in the multivariable models.

Overall, 32 (28.3%) of the 113 oncology trials were discontinued versus 70 (15.2%) of the 461 trials in other disease areas (table 2.2.3). Table S2.2.1 shows that this association is statistically significant after adjusting for the other trial characteristics (adjusted IRR 1.7; 95% CI 1.1-2.7). We also found that oncology trials were at statistically significant higher risk of efficacy-related discontinuation (adjusted IRR 2.5; 95% CI 1.2-5.1, table S2.2.2).

DISCUSSION

In our study, we showed that a substantial proportion (18%) of all clinical drug trials was discontinued before the planned end of recruitment and/or end of data collection. The proportion of discontinuation is within the range identified by previous studies of 11% - 45% [10,11,15-20]. Differences may be explained by different selection criteria, as previous studies also included non-drug trials, only randomized trials [10], or selection of exclusively oncology trials [11]. Further reasons for the varying results may be the dependence on registries, publications or questionnaires instead of IRB-files [16-18], or chance. Furthermore, our results show that the problem of poor recruitment remains of concern for in particular (but not only limited to) investigator-initiated trials. Recruitment estimations can be over-optimistic and should therefore be justified in the protocol. When in the trial protocol strict in- and exclusion criteria are given, investigators should provide data indicating that recruiting the needed number of participants from this population is feasible within the planned time. Literature and pilot research could for example identify whether sufficient candidate participants fulfilling the trial population criteria are willing to participate [21,22].

The percentage of discontinuations for futility and superiority reasons is consistent with the findings of Kasenda et al [10]. Discontinuation of a clinical trial after a well-designed interim analysis is not a failure. A research question can be answered by conducting the interim analysis at the right time, applying adequate stopping rules for statistical significance, and under supervision of an independent and skilled data monitoring committee. These aspects of the interim analysis should be described in the trial protocol. If the interim analysis is not described appropriately in the protocol, scientific objectivity is at risk to be preceded by personal or commercial motivations, for example through p-hacking[23]. Of the efficacy-related discontinuations in our study, two-thirds described at least a responsible DMC, the moment of the interim analysis in the trial, or the used stopping rule. However, only one-third described these three essential aspects of an adequate procedure for an interim analysis [8] in the trial protocol. The proportion of trials with at least some planning in the protocol in our study is considerably higher compared to the one-third found by Stegert et al.[24]. However, efficacy-related discontinuations are still often based on inadequately described procedures. The suggestion to improve trial protocols with regard to interim analyses is in particular, as our results show, for the industry-sponsored trials. Oncology trials were both at a statistically significant higher risk for all-cause discontinuation and for efficacy-related discontinuation. Possible explanations are the pressing need for effective therapies against various cancers[25] and the competitive drug market in oncology[26]. These reasons may be incentives to finish trials and act on their preliminary results. Our results show that these discontinuations are often not justified. The small number of discontinuations for superiority reasons in our study is contrary to the concerns expressed in the literature that this is a rising and questionable phenomenon [4,7,27,28]. It may be that these publications have led to a cautious attitude towards discontinuations due to interim analyses demonstrating superiority, diminishing its occurrence.

Six percent of the trials were discontinued due to recruitment failure, which is somewhat lower compared to the 10% found by Kasenda et al[10]. This figure was slightly lower in our study among randomized compared to non-randomized trials (22/418, 5.3% vs. 11/156, 7.1%, respectively), also when excluding the phase 1 trials (31/364, 7.8%). Another study previously found higher incidence of recruitment failure among randomized trials compared to non-randomized trials. The difference with our study may be explained by that they excluded crossover trials, or that they included relatively more phase 1 trials [29]. We replicated the finding that the risk of investigator-initiated trials to discontinue due to recruitment failure is more than twofold compared to industry-sponsored trials[10,30], although the small sample size prevented a statistically significant effect in our

multivariable model. Furthermore, we descriptively showed that phase 4 trials have a higher likelihood of discontinuation due to recruitment failure compared to other phases. Although the sample size was too low to test this association in multivariable analysis, it suggests that the motivation to recruit and/or to participate in a trial is limited after a drug also has become available in regular clinical practice. It also highlights the challenge of solving safety issues about newly approved drugs in the post-marketing phase [31].

A recent study showed that information about trial discontinuation is often not updated in trial registries [32]. In addition, the discontinued trials in our cohort remained significantly more often unpublished: 36% of the trials that were completed as planned remained unpublished versus 67% of the discontinued trials (manuscript submitted). Discontinued trials may be sometimes considered as failures and therefore as being not interesting or relevant to publish or disclose the details about. Nevertheless, transparency and traceability of such trials is important to prevent future failures for the same reasons.

The finding that only 14 trials were discontinued for safety reasons suggests that the likelihood of safety problems in drug trials is not very high (2.4%, table 2.2.1), and similar compared to other studies[10,24]. However, we did not have access to the individual trial safety data to further investigate this and thus the issue of safety is outside the scope of our study. Recent events show that the safety of trial participants remains of primary importance for investigators, sponsors and IRBs[33].

Discontinuations due to recruitment failure, financial reasons (90% of these were industry-sponsored), suspension after an inspection identified Good Clinical Practice issues, product manufacturing or regulatory issues, organizational issues, and after an interim analysis not or incompletely described in the protocol can be considered as being probably unjustified, but at least questionable for various reasons [2,4,5,20,21,25,27,28,34-38]. Together, these reasons sum up to 69 trials (12% of the cohort, table 2.2.1). Probably, a number of these discontinuations were due to unforeseeable misfortunes. Others may have been avoided if the conduct was preceded by a better trial protocol, planning, justification of sample size, and/or organization [21,24,34].

Based on our findings, we propose three recommendations for improvement of the conduct of clinical trials. These are relevant for all stakeholders. In particular, as the gatekeeper of clinical research, IRBs can play an important role in their implementation. The first recommendation is to include realistic sample size justifications and a critical assessment of the burden posed on trial participants. Future research should focus on how to measure the feasibility of recruitment numbers and timelines, enabling to reduce the rate of these trial failures. The

2

second recommendation is that the interim analysis plan in trial protocols should be improved [2,7]. Preventing discontinuations after unplanned interim analyses found futility or superiority can lead to less research waste, as trials completed as planned deliver information that is more useful and less influenced by chance [37]. The final recommendation is that IRBs should only approve trials with clear contracts stating that it is the responsibility of the sponsor to complete the trial and not allowing questionable reasons for discontinuation.

A strength of our study is that we included on a nationwide level all trials approved within the inclusion period, from 23 different IRBs. Therefore, the findings are both complete and can be considered as generalizable across the broad activity of clinical drug trials in the Netherlands. Our study adds geographic representativeness to the existing literature, as we were able to confirm largely the findings of trials reviewed by IRBs in Germany, Canada and Switzerland [10,24]. We had full access to the documents of the national competent authority and collaborated extensively with the local IRBs and investigators. Despite having access to a full cohort of drug trials, numbers in certain categories of potential determinants were small, with impacted our ability to obtain precise estimates in our multivariable models.

To conclude, one out of five clinical drug trials is discontinued before the planned trial end. Most of these discontinuations are related to recruitment failure, or interim analyses demonstrating futility. One out of eight clinical drug trials is discontinued for a questionable reason. IRBs should request more realistic recruitment targets. They should also request industry-sponsored multicenter trial applications to provide an adequate plan for an interim analysis in the trial protocol, including DMC oversight, the moment of the interim analysis in the trial progress, and the stopping rule that will be used.

SUPPLEMENTARY RESULTS

TABLE S2.2.1 Multivariable Poisson regression model for all-cause discontinuation versus completed as planned.

Total in model: 574	Completed as planned	Discontinued (all-cause)	Crude IRR (95% CI)	Adjusted IRR (95% CI)
	472 (82.2%)	102 (17.8%)		
Sample size				
Planned target sample size, median (IQR)	68 (24-314)	120 (40-392)	1.0 (1.0-1.0)	1.0 (1.0-1.0)
Sponsor				
Pharmaceutical industry (N = 352)	288 (81.8%)	64 (18.2%)	ref	ref
Investigator-initiated (N = 222)	184 (82.9%)	38 (17.1%)	0.9 (0.6-1.4)	1 (0.6-1.6)
Phase				
Phase 1 (N = 119)	108 (90.8%)	11 (9.2%)	ref	ref
Other phases (N = 455)	364 (80.0%)	91 (20.0%)	2.2 (1.2-4.0)	1.8 (0.9-3.7)
Centers				
Single center (N = 249)	219 (88.0%)	30 (12.0%)	ref	ref
Multicenter (N = 325)	253 (77.8%)	72 (22.2%)	1.8 (1.2-2.8)	1.5 (0.8-2.6)
Randomization				
Randomized trial (N = 418)	344 (82.3%)	74 (17.7%)	1.0 (0.7-1.6)	0.9 (0.6-1.5)
Non-randomized trial (N = 156)	128 (82.1%)	28 (17.9%)	ref	ref
Disease area				
Oncology (N = 113)	81 (71.7%)	32 (28.3%)	1.9 (1.2-2.8)	1.7 (1.1-2.7)
Other than oncology (N = 461)	391 (84.8%)	70 (15.2%)	ref	ref

Abbreviations: IRR = incidence rate ratio; CI = confidence interval; IQR = interquartile range.

TABLE S2.2.2 Multivariable Poisson regression model for efficacy-related discontinuation versus completed as planned in the subgroup of phase 2 and 3 trials.

Total in model: 260*	Completed as planned	Discontinued solely efficacy-related	Crude IRR (95% CI)	Adjusted IRR (95% CI)
	231 (88.8%)	29[†] (11.2%)		
Sample size				
Planned target sample size, median (IQR)	260 (63-658)	317 (83-649)	1.0 (1.0-1.0)	1.0 (1.0-1.0)
Disease area				
Oncology (N = 70)	56 (80.0%)	14 (20.0%)	2.5 (1.2-5.2)	2.5 (1.2-5.1)
Other than oncology (N = 188)	175 (93.1%)	13 (6.9%)	ref	ref

Abbreviations: IQR = interquartile range; IRR = incidence rate ratio; CI = confidence interval. *****This model was restricted to the phase 2 and phase 3 trials of the cohort. The control group were phase 2 and phase 3 trials that were completed as planned. [†] One efficacy-related discontinuation was a phase 1 trial (table 2.2.3) and is therefore not included in this model.

TABLE S2.2.3 Multivariable Poisson regression model for discontinuation due to recruitment failure versus completed as planned in the subgroup of non-phase 1 trials.

Total in model: 395*	Completed as planned	Discontinued due to recruitment failure	Crude IRR (95% CI)	Adjusted IRR (95% CI)
	364 (92.2%)	31[†] (7.8%)		
Sample size				
Planned target sample size, median (IQR)	107 (33-494)	80 (25-200)	0.9 (0.8-1.0)	0.9 (0.8-1.1)
Sponsor				
Pharmaceutical industry (N = 202)	193 (95.5%)	9 (4.5%)	ref	ref
Investigator-initiated (N = 193)	171 (88.6%)	22 (11.4%)	2.6 (1.2-5.6)	2.0 (0.9-4.6)

Abbreviations: IQR = interquartile range; IRR = incidence rate ratio; CI = confidence interval. *This model was restricted to the phase 2, 3, 4, and other than phase 1-4 trials of the cohort. The control group was non-phase 1 trials that were completed as planned. [†]Two discontinuations due to recruitment failure were phase 1 trials (table 2.2.3) and are therefore not included in this model.

REFERENCES

1.   Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. *N Engl J Med.* Aug 10 1989;321(6):406-412.

2.   Pocock SJ. When (not) to stop a clinical trial for benefit. *JAMA.* Nov 2 2005;294(17):2228-2230.

3.   van der Graaf R, van Delden JJ. Equipoise should be amended, not abandoned. *Clin Trials.* Aug 2011;8(4):408-416.

4.   Bassler D, Montori VM, Briel M, Glasziou P, Guyatt G. Early stopping of randomized clinical trials for overt efficacy is problematic. *Journal of clinical epidemiology.* Mar 2008;61(3):241-246.

5.   Pocock S, White I. Trials stopped early: too good to be true? *Lancet.* Mar 20 1999;353(9157):943-944.

6.   Pocock SJ, Stone GW. The Primary Outcome Is Positive - Is That Good Enough? *N Engl J Med.* Sep 8 2016;375(10):971-979.

7.   Montori VM, Devereaux PJ, Adhikari NK, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA.* Nov 2 2005;294(17):2203-2209.

8.   ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group. *Stat Med.* Aug 15 1999;18(15):1905-1942.

9.   Kesselheim AS, Avorn J. The most transformative drugs of the past 25 years: a survey of physicians. *Nat Rev Drug Discov.* Jun 2013;12(6):425-431.

10. Kasenda B, von Elm E, You J, et al. Prevalence, characteristics, and publication of discontinued randomized trials. *JAMA.* Mar 12 2014;311(10):1045-1051.

11. Schroen AT, Petroni GR, Wang H, et al. Preliminary evaluation of factors associated with premature trial closure and feasibility of accrual benchmarks in phase III oncology trials. *Clin Trials.* Aug 2010;7(4):312-321.

12. van den Bogert CA, Souverein PC, Brekelmans CT, et al. Occurrence and determinants of selective reporting of clinical drug trials: design of an inception cohort study. *BMJ Open.* 2015;5(7):e007827.

13. van den Bogert CA, Souverein PC, Brekelmans CT, et al. Non-Publication Is Common among Phase 1, Single-Center, Not Prospectively Registered, or Early Terminated Clinical Drug Trials. *PLoS One.* 2016;11(12):e0167709.

14. Arshad A, Arkwright PD. Status of healthcare studies submitted to UK research ethics committees for approval in 2004-5. *J Med Ethics.* May 2008;34(5):393-395.

15. Pich J, Carne X, Arnaiz JA, Gomez B, Trilla A, Rodes J. Role of a research ethics committee in follow-up and publication of results. *Lancet.* Mar 22 2003;361(9362):1015-1016.

16. Decullier E, Lheritier V, Chapuis F. Fate of biomedical research protocols and publication bias in France: retrospective cohort study. *BMJ.* Jul 2 2005;331(7507):19.

17. Easterbrook PJ, Matthews DR. Fate of research studies. *J R Soc Med.* Feb 1992;85(2):71-76.

18. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ.* Sep 13 1997;315(7109):640-645.

19. von Elm E, Rollin A, Blumle A, Huwiler K, Witschi M, Egger M. Publication and non-publication of clinical trials: longitudinal study of applications submitted to a research ethics committee. *Swiss Med Wkly.* Apr 5 2008;138(13-14):197-203.

20. Carlisle B, Kimmelman J, Ramsay T, MacKinnon N. Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials. *Clin Trials.* Feb 2015;12(1):77-83.

21. Briel M, Olu KK, von Elm E, et al. A systematic review of discontinued trials suggested that most reasons for recruitment failure were preventable. *Journal of clinical epidemiology.* Aug 3 2016.

22. Thabane L, Ma J, Chu R, et al. A tutorial on pilot studies: the what, why and how. *BMC Med Res Methodol.* 2010;10:1.

23. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-hacking in science. *PLoS Biol.* Mar 2015;13(3):e1002106.

24. Stegert M, Kasenda B, von Elm E, et al. An analysis of protocols and publications suggested that most discontinuations of clinical trials were not based on preplanned interim analyses or stopping rules. *J Clin Epidemiol.* Jan 2016;69:152-160.

25. Trotta F, Apolone G, Garattini S, Tafuri G. Stopping a trial early in oncology: for patients or for industry? *Ann Oncol.* Jul 2008;19(7):1347-1353.

26. Nawaz K, Webster RM. The non-small-cell lung cancer drug market. *Nat Rev Drug Discov.* Apr 2016;15(4):229-230.

27. Casazza G, Casella F. Can we trust in trials stopped early for benefit? *Intern Emerg Med.* Dec 2012;7(6):559-561.

28. Heffner AC, Milbrandt EB, Venkataraman R. Trials stopped early for benefit? Not so fast! *Crit Care.* 2007;11(1):305.

29. Blumle A, Schandelmaier S, Oeller P, et al. Premature Discontinuation of Prospective Clinical Studies Approved by a Research Ethics Committee - A Comparison of Randomised and Non-Randomised Studies. *PLoS One.* 2016;11(10):e0165605.

30. Damen L, van Agt F, de Boo T, Huysmans F. Terminating clinical trials without sufficient subjects. *J Med Ethics.* Jul 2012;38(7):413-416.

31. Vermeer NS, Duijnhoven RG, Straus SM, et al. Risk management plans as a tool for proactive pharmacovigilance: a cohort study of newly approved drugs in Europe. *Clin Pharmacol Ther.* Dec 2014;96(6):723-731.

32. Alturki R, Schandelmaier S, Olu KK, et al. Premature trial discontinuation often not accurately reflected in registries: comparison of registry records with publications. *Journal of Clinical Epidemiology.* Sep 7 2016.

33. Moore N. Lessons from the fatal French study BIA-10-2474. *BMJ.* 2016;353:i2727.

34. Lievre M, Menard J, Bruckert E, et al. Premature discontinuation of clinical trial for reasons not related to efficacy, safety, or feasibility. *BMJ.* Mar 10 2001;322(7286):603-605.

35. Evans S, Pocock S. Societal responsibilities of clinical trial sponsors. Lack of commercial pay off is not a legitimate reason for stopping a trial. *BMJ.* Mar 10 2001;322(7286):569-570.

36. Pocock SJ. Size of cancer clinical trials and stopping rules. *Br J Cancer.* Dec 1978;38(6):757-766.

37. Pocock SJ. When to stop a clinical trial. *BMJ.* Jul 25 1992;305(6847):235-240.

38. Grant AM, Altman DG, Babiker AB, et al. Issues in data monitoring and interim analysis of trials. *Health Technol Assess.* Mar 2005;9(7):1-238, iii-iv.

2

Chapter 2.3

Non-publication is common among phase 1, single-center, not prospectively registered, or early terminated clinical drug trials

Cornelis A. van den Bogert; Patrick C. Souverein; Cecile T.M. Brekelmans; Susan W.J. Janssen; Gerard H. Koëter; Hubert G.M. Leufkens; Lex M. Bouter

ABSTRACT

Objectives

The objective of this study was to investigate the occurrence and determinants of non-publication of clinical drug trials in the Netherlands.

Methods

All clinical drug trials reviewed by the 28 Institutional Review Boards (IRBs) in the Netherlands in 2007 were followed-up from approval to publication. Candidate determinants were trial characteristics. The main outcome was publication as peer-reviewed article. The percentage of trials that were published, crude and adjusted odds ratio (OR), and 95% confidence interval (CI) were used to quantify the associations between determinants and publication.

Results

Of the 574 analyzed trials, 334 (58%) were published as peer-reviewed article. The following determinants were statistically significant associated with publication: phase 2 (60% published; adjusted OR 2.6, 95% CI 1.1-5.9), phase 3 (73% published; adjusted OR 4.1, 95% CI 1.7-10.0), and trials not belonging to phase 1-4 (60% published; adjusted OR 3.2, 95% CI 1.5 to 6.5) compared to phase 1 trials (35% published); trials with a company or investigator as applicant (63% published) compared to trials with a Contract Research Organization (CRO) as applicant (50% published; adjusted OR 1.7; 95% CI 1.1-2.8); and multicenter trials also conducted in other EU countries (68% published; adjusted OR 2.2, 95% CI 1.1-4.4) or also outside the European Union (72% published; adjusted OR 2.0, 95% CI 1.0-4.0) compared to single-center trials (45% published). Trials that were not prospectively registered (48% published) had a lower likelihood of publication compared to prospectively registered trials (75% published; adjusted OR 0.5, 95% CI 0.3-0.8), as well as trials that were terminated early (33% published) compared to trials that were completed as planned (64% published; adjusted OR 0.2, 95% CI 0.1-0.3).

Conclusions

The non-publication rate of clinical trials seems to have improved compared to previous inception cohorts, but is still far from optimal, in particular among phase 1, single-center, not prospectively registered, and early terminated trials.

2

INTRODUCTION

Since decades, non-publication of trial results has been a major concern in clinical research, as non-publication causes research waste[1,2], and can bias evidence-based treatment guidelines and clinical decision making[3-5]. Research waste was defined by Chalmers and Glasziou as avoidable waste of investments in research due to inadequately producing and reporting, non-publication being one of its four stages[1]. In 2009, the magnitude of research waste in clinical research was estimated at 85%[1]. Moreover, non-publication is unethical because the burdens and risks imposed on study participants do not contribute to the body of knowledge.

The implications of waste and bias in research caused by non-publication[3,6-18] strengthens the view that all clinical trials must be published[19-23]. Previous studies specifically focused on publication of randomized controlled trials (RCTs)[24], covered only trials within one medical specialty[25], examined a limited selection of determinants, or used incomplete trial cohorts depending on public registrations[26,27] or interview response rates[10]. As a result, there is limited data on the occurrence of non-publication and its determinants that is both recent and complete. The most well-known determinant for non-publication is having a 'negative' outcome[28]. However, other reasons for non-publication have been proposed (e.g. rejection by editors or influence of the sponsor[29]). Investigating determinants of non-publication can identify and provide specific solutions for areas where the problem of research waste and bias is most persistent. Therefore, the aim of our study was to investigate the occurrence and determinants of non-publication of clinical drug trials in a countrywide inception cohort of clinical drug trials.

METHODS

The design of our study and the characteristics of the included trials have been published elsewhere[30]. In short, the inception cohort consisted of all clinical drug trials reviewed by IRBs in the Netherlands between 1 January and 31 December 2007. We used ToetsingOnline[31], the database of the competent authority of the Netherlands (the Central Committee on Research Involving Human Subjects, abbreviated in Dutch as CCMO), the only source containing a complete record of all trials that underwent IRB-review, to identify the cohort, the determinants, and the stages of progress of the included trials. In addition, we searched the trial registries clinicaltrials.gov and ISRCTN for the candidate determinant prospective registration, and for the availability of trial results in public reg-

istries. We originally defined prospective registration as registration before the first patient is recruited[30]. Because start-of-trial dates were missing in the database, we changed the definition of prospective registration to registration within one month of IRB-approval. In our experience, most trials start recruitment later than one month after IRB-approval, so this threshold classified more not prospectively registered trials as prospectively registered than vice versa. Sensitivity analyses were performed using two less strict thresholds of prospective registration: registration within 1 year of IRB-approval, and registration at any moment.
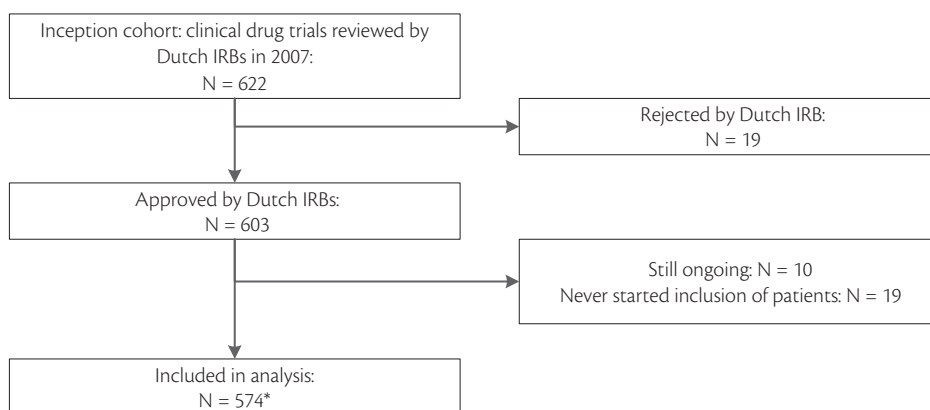

FIGURE 2.3.1 Selection of the samples for the analysis of the primary outcome, determinant analysis and protocol evaluation, starting with the inception cohort of all IRB-reviewed trials in 2007. *The end-of-trial form was missing of 186 of the 574 trials that were included in the analysis. Principal investigators of 73 of these trials responded to our questionnaire, completing the information on the end-of-trial. From the remaining 113, of 87 trials we found other documents than the end-of-trial form indicating that the trial had started (for example, emails from the IRB or amendments), or we found that the trial was published. Of 26 trials included in the analysis, we had no follow-up information. The 113 trials with missing information about completion were assumed to be completed as planned.

The search algorithm for publications used the platforms Pubmed, Embase and Google Scholar. More details are reported in the protocol[30]. We conducted the final search for publication and availability of results in January and February 2016. Therefore, the follow-up since IRB-approval was 8 years at minimum, and 9 at maximum. Questionnaires were e-mailed to the principal investigators (PIs) of the trials, asking for reasons for non-publication. If the PI had left the company or the hospital that conducted the trial, we tried to contact the PI at his current affiliation, or otherwise we attempted to contact colleagues of the PI that were involved in the same trial. After identification of the right person, at maximum two reminders were sent. The Dutch accredited IRBs were asked for permission to send the questionnaire to the PIs. All IRBs consented and provided

a signed letter of endorsement, which we attached to the questionnaire. The list of 23 Dutch accredited IRBs can be found on the website of the CCMO[32]. More information about the questionnaires is provided in appendix 2.2, including copies of the questionnaire templates (these can also be freely downloaded through the publication of this chapter in PLoS One[33]).

Candidate determinants were trial characteristics that the PI filled out on a form at the time of submission of the trial application for IRB-review. This form is mandatory and identical for all IRBs in the Netherlands. Prospective registration on the registries of clinicaltrials.gov or ISRCTN, and whether the trial was completed as planned or terminated early were also candidate determinants.

To be consistent with the literature referred to above, and for the purpose of linguistic clarity, we used publication as an outcome rather than non-publication. A publication was defined as a peer-reviewed article (i.e. the reciprocal of non-publication). Percentages of published trials were calculated for each of the determinant categories. Logistic regression was used to calculate crude and adjusted odds ratios (ORs) and 95% CIs for the association between determinants and publication. The final multivariable model included determinants that were retained after backward stepwise elimination based on the likelihood ratio, using $p>0.2$ as elimination rule. The original published study protocol prescribed Cox-regression for multivariable analysis instead of logistic regression[30]. However, the hazard ratios of determinants were not proportional during the observation period. Moreover, the end-of-trial dates were missing for 186 trials. Therefore, the date of IRB-approval was used as the starting point of follow-up, instead of the end-of-trial date prescribed by the protocol[30]. Because we were unable to control for the duration of the trials, interpretation of the hazard ratio would therefore be challenging and we decided to use logistic regression instead. The Kaplan Meier analysis was used to visualize the cohort from its starting point (date of IRB-review) until the endpoint (publication or non-publication), stratified by trial phase, one of the key determinants which also discriminates between longer- and shorter-during trials[34].

We also stratified by oncology versus other disease areas (pre-specified in the protocol), and further stratified oncology trials by phase 1 trials versus other phase trials (post-hoc). Oncology phase 1 trials differ from other disease area phase 1 trials in that oncology phase 1 trials are usually restricted to patients, while most other disease areas include healthy volunteers[35].

In a second post hoc analysis, we investigated the association of the direction of results and publication. We categorized the direction of conclusions as positive, negative or descriptive. This categorization was based on the conclusion paragraph of the publication (e.g. the investigated treatment was superior,

TABLE 2.3.1 Frequencies and publication percentages of candidate determinants.

	N in analysis (% published)
All trials included in the analysis	574 (58.2%)
Sponsor	
Pharmaceutical industry	352 (59.1%)
Investigator (industry (co-)funded)	71 (52.1%)
Investigator (no industry funding involved)	151 (58.9%)
Phase	
Phase 1	119 (34.5%)
Phase 2	130 (60.0%)
Phase 3	172 (72.7%)
Phase 4	57 (56.1%)
Other than phase 1-4*	96 (60.4%)
Applicant	
Contract research organization	214 (50.0%)
Investigator or company	360 (63.1%)
Centers	
Single center	249 (45.4%)
Multi center only in the Netherlands	54 (53.7%)
Multi center in the Netherlands and the EU	82 (68.3%)
Multi center in the Netherlands and outside the EU	189 (72.0%)
Therapeutic effect expected	
Therapeutic effect expected[†]	356 (64.6%)
No therapeutic effect expected	218 (47.7%)
Type of trial	
Intervention	517 (59.8%)
Invasive observational[‡]	45 (42.2%)
Non-invasive observational	12 (50.0%)
Approval status of drug(s) in trial	
Unapproved drug(s) in trial	306 (54.6%)
All drugs in trial approved, studied outside approved indication	147 (65.3%)
All drugs in trial approved and studied within approved indication	121 (58.7%)
Drug type	
Regular medicinal product	549 (57.7%)
Special drug category involved[§]	25 (68.0%)
Participant category	
≥18 years old and mentally capacitated	532 (58.6%)
<18 years old and/or mentally incapacitated	42 (52.4%)

TABLE 2.3.1 Frequencies and publication percentages of candidate determinants. (continued)

	N in analysis (% published)
Disease area	
Oncology	113 (66.4%)
Other disease areas	461 (56.2%)
Prospective registration[||]	
Prospectively registered	215 (74.9%)
Not (prospectively) registered	359 (48.2%)
Completion	
Completed as planned	472 (63.6%)
Terminated early	102 (33.3%)

*Studies not primarily intended to provide information about the drug, nor conducted within the context of a drug development program. [†]Trials were regarded as therapeutic if it is reasonable to assume that participation will be of direct clinical benefit to the subject. [‡]In observational trials, the investigator does not seek to change the observed situation, but simply to describe and record it as accurately as possible. Invasive procedures concern the penetration of the skin or mucosa with the aid of instruments, X-rays or magnetic resonance, or the introduction of an instrument into the body, or psychologically invasive observational research, involving the experimental creation of an unaccustomed situation that may give rise to negative emotions in the subject. [§]Vaccine, radiopharmaceutical, somatic cell therapy, antisense oligonucleotide. [\|\|] Prospective registration was defined as registration of the trial at www.clinicaltrials.gov or www.isrctn.com, at latest one month after IRB-approval.

equivalent, and/or safer than the comparator), and for the unpublished trials on the primary outcome measurement reported in the registry (positive if the primary outcome was in favor of the investigated treatment, negative if not, and descriptive if no statistical test was provided in the registry). All data analyses were performed in IBM SPSS Statistics, version 23.

RESULTS

Of the 622 trials reviewed by the Dutch IRBs, 19 (3.0%) were rejected, and after obtaining IRB-approval, another 19 trials never started the inclusion of patients (figure 2.3.1). Thus, before any patients were included, 6% of the trials had reached their final stage of progress. Of the 574 trials that started, 334 trials (58.2%) were published within the observation period of 8-9 years after IRB-approval.

Table 2.3.1 shows all candidate determinants and the percentages of publication for each determinant category. Nine of these candidate determinants were in-cluded in the multivariable logistic regression model (table 2.3.2). In this model, phase 2 (adjusted OR 2.6; 95% CI 1.1-5.9), 3 (adjusted OR 4.1; 95% CI 1.7-10.0)

and other-phase trials (adjusted OR 3.2; 95% CI 1.5-6.5) had a significantly higher likelihood of publication compared to phase 1 trials. Trials of which the investigator or company was the applicant had a significantly higher likelihood of publication compared to trials of which a contract research organization was the applicant (adjusted OR 1.7; 95% CI 1.1-2.8). Furthermore, international multicenter trials within the EU (adjusted OR 2.2; 95% CI 1.1-4.4) or also outside the EU (adjusted OR 2.0; 95% CI 1.0-4.0) were more likely published than single center trials. Invasive observational trials had a lower likelihood of publication compared to intervention trials (adjusted OR 0.4; 95% CI 0.2-0.9). Trials that were not prospectively registered had a lower likelihood of publication compared to prospectively registered trials (adjusted OR 0.5; 95% CI 0.3-0.8). Sensitivity analyses showed that the magnitude of this association increased if the threshold of prospective registration was changed to registration within one year of IRB-approval, or to registration at any moment (data not shown). Finally, trials that were terminated early had a substantially lower likelihood of publication compared to trials that were

Visually, the Kaplan Meier curves of all phases seemed to approach their plateau after 8-9 years of follow-up since IRB-approval (figure 2.3.2). The overall median time to publication since IRB-approval was 53 months (interquartile range (IQR) 39-65) and was not different between the trial phases.

Overall, non-oncology trials had a lower likelihood of publication compared to oncology trials; however, this association was not significant in the multivariable analysis (table 2.3.2, adjusted OR 0.7, 95% CI 0.4-1.1, supplementary results figure S2.3.1). No significant difference was observed in the median time to publication between other disease area and oncology trials (median time to publication 52 months (IQR 41-69) vs. 57 months (IQR 39-63), respectively). Post-hoc analysis showed that only 28 out of 100 (28%) other disease area phase 1 trials were published, which was significantly lower compared to the 13 out of 19 (68%) published oncology phase 1 trials (OR 0.2, 95% CI 0.1-0.5; supplementary results figure S2.3.2). Among other phases, we observed no difference in publication of other disease area and oncology trials (64% vs. 66%, respectively; supplementary results figure S2.3.3).

TABLE 2.3.2 Associations between determinants and publication, expressed as crude and adjusted odds ratios (OR), and 95% confidence intervals (CI) of the crude and adjusted ORs.

Determinants	Crude OR (95% CI)	Adjusted OR (95% CI)
Phase		
Phase 1	ref	ref
Phase 2	2.9 (1.7-4.8)	2.6 (1.1-5.9)
Phase 3	5.1 (3.1-8.4)	4.1 (1.7-10.0)
Phase 4	2.4 (1.3-4.6)	2.4 (0.9-6.3)
Other than phase 1-4	2.9 (1.7-5.1)	3.2 (1.5-6.5)
Applicant		
Contract research organization	ref	ref
Investigator or company	1.7 (1.2-2.4)	1.7 (1.1-2.8)
Centers		
Single center	ref	ref
Multicenter only in the Netherlands	1.4 (0.8-2.5)	1.2 (0.6-2.4)
Multicenter in the Netherlands and the EU	2.6 (1.5-4.4)	2.2 (1.1-4.4)
Multicenter in Netherlands and outside EU	3.1 (2.1-4.6)	2.0 (1.0-4.0)
Therapeutic effect expected		
Therapeutic effect expected	ref	ref
No therapeutic effect expected	0.5 (0.4-0.7)	1.7 (0.9-3.3)
Type of trial		
Intervention	ref	ref
Invasive observational	1.5 (0.5-4.7)	0.4 (0.2-0.9)
Non-invasive observational	0.7 (0.2-2.6)	0.9 (0.3-3.2)
Participant category		
≥18 years old and able to provide consent	ref	ref
<18 years old and/or unable to provide consent	0.8 (0.4-1.5)	0.5 (0.2-1.0)
Disease area		
Oncology	ref	ref
Other disease areas	0.7 (0.4-1.0)	0.7 (0.4-1.1)
Prospective registration		
Prospectively registered	ref	ref
Not (prospectively) registered	0.3 (0.2-0.5)	0.5 (0.3-0.8)
Completion		
Completed as planned	Ref	ref
Terminated early	0.3 (0.2-0.5)	0.2 (0.1-0.3)

2

Substantially more published trials (113/334, 34%) had also uploaded a summary of results in the clinicaltrials.gov or ISRCTN registries compared to the unpublished trials (23/240, 10%). Post hoc analyses showed that of the published trials, 42% of the direction of conclusions was positive, 19% was negative, and 39% were descriptive. Of the unpublished trials that reported results in their registry, 5 (22%) trials reported a positive primary outcome, 2 (9%) reported a negative primary outcome and 16 (70%) were descriptive or missing (primarily due to missing statistical information that was needed to infer a direction of the results).
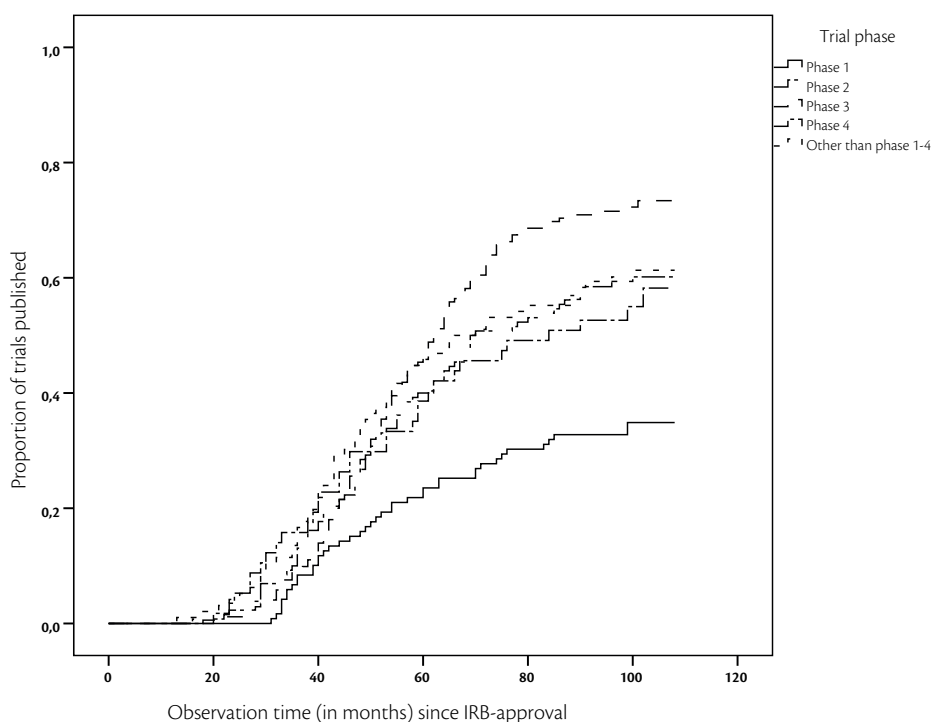

FIGURE 2.3.2 Kaplan Meier analysis of the publication rates of trial phases. IRB = Institutional Review Board.

The principal investigators of only 55 of the 240 (23%) unpublished trials responded to the questionnaire and provided the reason(s) for non-publication (supplementary results table S2.3.1). The most important reason for non-publication among the responders was that the investigators had other priorities than to write a manuscript (18.2%). Other reasons included no statistically significant or clinically relevant results (14.5%), the manuscript was rejected by a journal (12.7%), the article was not finished yet (10.9%), and the study was underpowered due to poor inclusion of participants (10.9%).

DISCUSSION

Of the clinical drug trials approved by the Dutch IRBs in 2007, 42% had not been published as a peer-reviewed article by January/February 2016. The publication rate approximated their plateau at the time of our final search, suggesting that only a few more publications can be expected. The observed publication rate of 42% is relatively high compared with other studies investigating older cohorts[3,6-9,11-14,16,36,37]. This suggests that the publication rate of clinical trials has somewhat improved, but is still far from ideal. In particular, the publication percentage of the phase 3 trials (mainly RCTs) in our cohort (73%) was higher compared to previous cohorts investigating RCTs (overall, 54% published)[16,38]. Other recent research also supports that publication of phase 3 trials has been improved[17]. Therefore, the regularly mentioned number of 50% non-publication[39] probably needs to be updated with regard to the phase 3 trials. Awareness-raising public campaigns[40], incorporation of publication requirements in clinical trial legislation[41] and advocacy by influential organizations[20] over the past decade may have contributed to this improvement. However, it is uncertain whether the identified publications have adequately reported all relevant aspects of the trials[42]. We are investigating this in the next phase of our cohort study[30].

The implicated research waste is considerable. Starting with the inception cohort of 622 IRB-reviewed trials, at least 140 (23%) failed to be completed as planned (figure 2.3.1 and table 2.3.1). If we consider the published trials and the trials that are still running as not (yet) wasted, waste is implicated in 50% of the trials. This percentage should not be compared to the research waste estimate of 85% (of which 50% was due to non-publication) suggested by Chalmers and Glasziou[1], as we did not factor in research waste due to a poor design, conduct, data analysis, and selective reporting within the publications. Some waste is probably unavoidable (for example, trials sometimes are terminated early for ethical reasons). However, the need for better solutions is urgent considered the large public and private investments involved in the unpublished trials. Furthermore, 42% non-publication implies that publication bias in clinical drug trials is likely still substantial, despite many years of attention to this topic [43].

A limitation of our study was that we did not include the direction, magnitude and statistical significance of the trial results as determinants in our analysis. Previous studies included this determinant[10,15], by interviewing the PIs[10], or using trial reports submitted to the IRB[15]. However, this approach excludes trials of which no such data is available, potentially introducing selection bias. This would have excluded 113 of the 240 (77%) unpublished trials from our cohort. Furthermore, it is questionable how objective investigators can judge the direc-

tion of results of their own research[44], and definitions of 'positive' and 'negative' results are heterogeneous[28]. Despite the attached endorsement letters from the local IRBs, the response rate to our questionnaire was low. Among the responders, only 14.5% of the PIs reported that lack of significance or relevance of the results were a reason for non-publication. Having other priorities was the most common reason. Rejection by a journal was also among the most common reasons for non-publication. Both these reasons have been reported previously in the literature[16,45]. The post hoc analysis of the results of the unpublished trials that were uploaded in their registry demonstrated that these results sections are often incomplete and provide therefore little information on the influence of the direction of the results on the likelihood of publication. Furthermore, this finding suggests in line with other studies that uploading results in trial registries should be done more often, and that the quality of these results uploads needs improvement[46,47].

The publication rate of phase 1 trials was substantially lower compared to other phases. This has been shown before[8]. However, the percentage of phase 1 trials that was published in our cohort was substantially higher (35%) than the previous study (17%)[8], suggesting that progress has also been made in the field of phase 1 trials, but still not sufficient. Publication of phase 1 trials may be considered less interesting because their direct impact for clinical practice is limited when the drug is still far from marketing approval. Yet, phase 1 trials are an important source for the clinical pharmacology of drugs. Furthermore, data from previous phase 1 trials on similar drugs is essential in determining the risk of phase 1 (first in man) trials upfront[48]. Increasing transparency in general in this field of clinical research should be high on the agenda of regulators and the pharmaceutical industry, as emphasized by the slow release of information after the recent tragic events in a phase 1 trial in France[49].

Our post hoc finding that oncology phase 1 trials are more likely to be published than phase 1 trials in other disease areas suggests that inclusion of patients who are typically very ill[50] may positively influence publication of phase 1 trials. Or, argued differently, oncology phase 1 trials are in fact phase 2 trials, as phase 2 trials in most other disease areas are usually the 'first-in-patient' trials. The publication percentage of oncology phase 1 trials in our cohort was indeed similar to that of the phase 2 trials (68% and 60%, respectively).

The lower likelihood of publication of single center trials compared to multi-center trials has been shown in previous research[10]. In our cohort, this trend was visible, but only statistically significant for multicenter trials conducted also outside the Netherlands. Opportunities for increasing the incentive to publish exist at the level of the trial center. Publication metrics (including, but not limited to

the number of trials published divided by the total number of trials conducted) should be reported on the center-website as well as the website of the local IRB for all trials conducted in the center[51]. Transparency about the local publication practices may stimulate stakeholders to require publication of all trials.

Invasive observational trials had a lower likelihood to be published compared to intervention trials. This association was not observed between observational non-invasive trials and intervention trials. Findings by other studies regarding this determinant are inconsistent[52] and the poor precision makes this determinant difficult to interpret.

We found that prospective registration in a trial registry was associated with publication. The idea of prospective registration of all trials was proposed many years ago[4], but in our cohort, only 37% of the trials were prospectively registered. The sensitivity analyses showed that the significant association with publication remained when using the less strict definition of prospective as registration within 1 year of IRB-approval. Since 2007, prospective registration has become increasingly mandatory, and higher registration rates have been reported[53]. However, given the changes in the requirements for prospective registration since the inception of this cohort, higher publication rates cannot be predicted from this rise in prospective registration. Furthermore, there is no evidence that registries in their current state can adequately replace journal articles as the primary source for clinical guidelines, decision making and designing future trials. Until the issues with registries, such as completeness and quality of uploads of trial results, are solved, the peer-reviewed journal article remains the golden standard for reporting the results of clinical trials, and all clinical trials should be published as such.

To conclude, our study shows a non-publication rate of clinical trials of 42%, which seems to be an improvement compared to previous inception cohorts, but is still far from optimal. Determinants of non-publication are early termination, no prospective registration, phase 1, and single center. Nevertheless, considerable waste is implicated, and the likelihood of publication bias is high.

SUPPLEMENTARY RESULTS

TABLE S2.3.1 Reported reasons for non-publication (responses were obtained for 55/240 unpublished trials).

| Reason for non-publication	Frequency reason was reported	Percentage of the 55 responding PIs
Other priorities	10	18.2%
Results not clinically relevant	7	12.7%
Manuscript rejected by journal	7	12.7%
Article/analysis is not finished yet	6	10.9%
Low number of participants, therefore underpowered	6	10.9%
Discontinued development of the drug	5	9.1%
Study was preliminary terminated	4	7.3%
Results not statistically significant	3	5.5%
Study was presented on conference	3	5.5%
Study only intended for development of drug	3	5.5%
Results may be published after drug approval	2	3.6%
No reason provided/known	2	3.6%
Investigators felt not responsible to publish	2	3.6%
Journal space restrictions	1	1.8%
Sponsor decision	1	1.8%
Results were not spectacular	1	1.8%
Drug development was transferred to other company	1	1.8%
Not included in trial register	1	1.8%
Phase 1 study	1	1.8%
Validity of data questioned by health authorities	1	1.8%
Only reported internally	1	1.8%
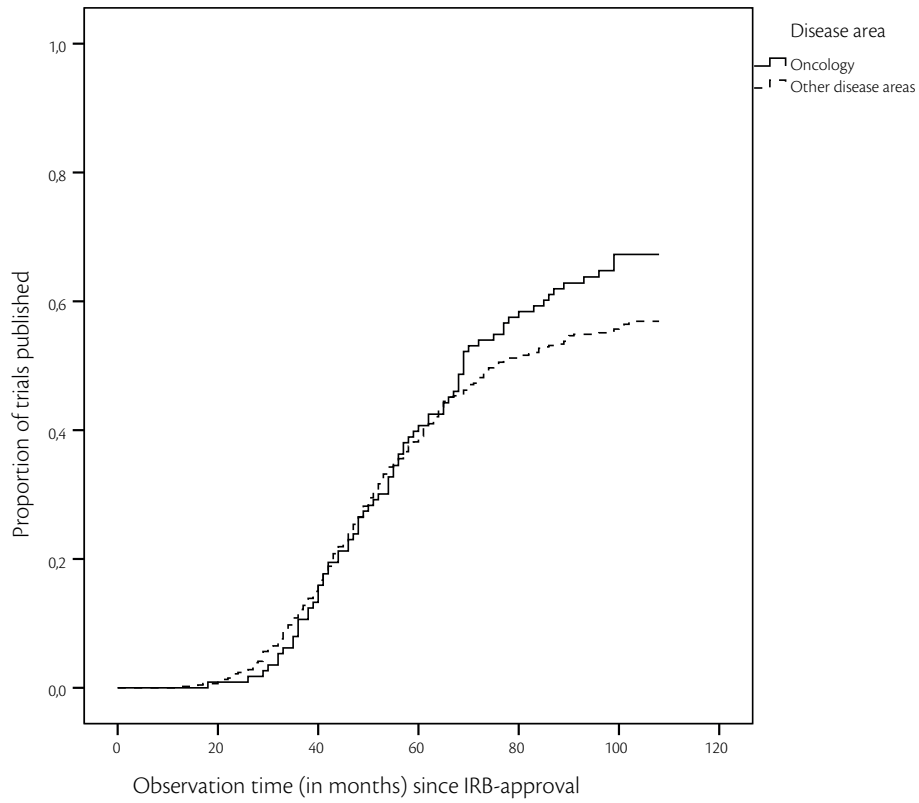
FIGURE S2.3.1 Kaplan Meier analysis of the publication rates of all trials, stratified by oncology versus all other disease areas.
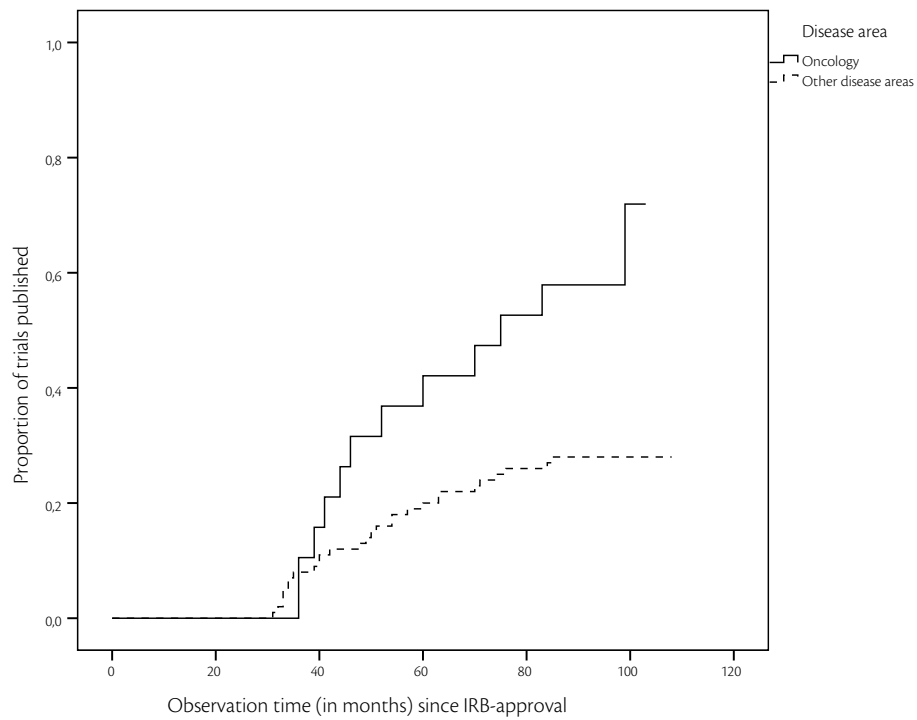
FIGURE S2.3.2 Kaplan Meier analysis of the publication rates of phase 1 trials, stratified by oncology versus all other disease areas.
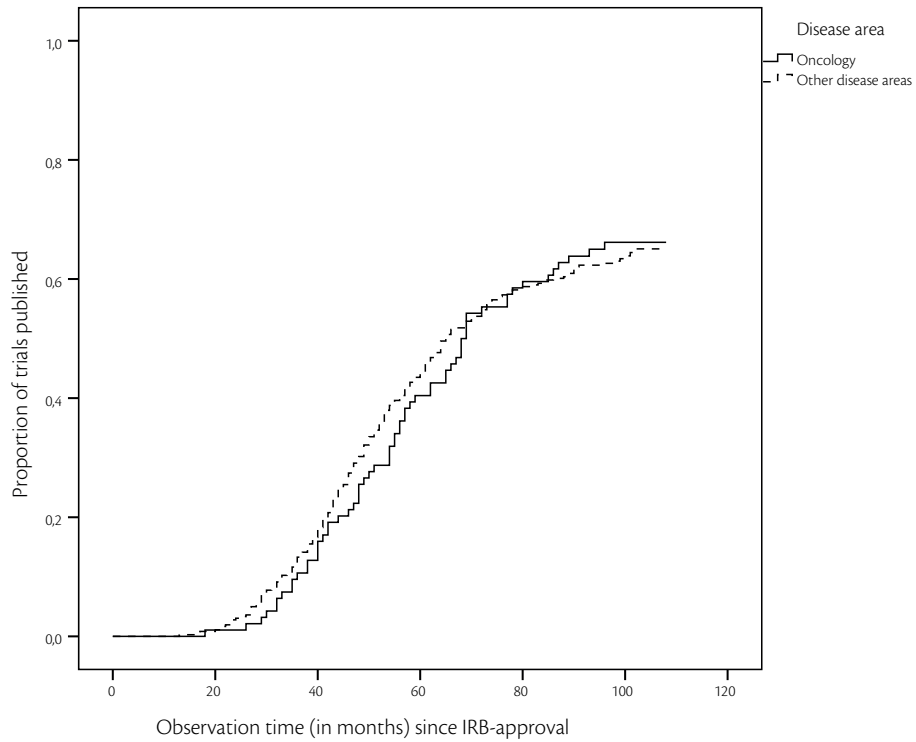
FIGURE S2.3.3 Kaplan Meier analysis of the publication rates of all trials except phase 1, stratified by oncology versus all other disease areas.

REFERENCES

1. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. Jul 4 2009;374(9683):86-89.

2. Ioannidis JP. Clinical trials: what a waste. *BMJ*. 2014;349:g7089.

3. Bardy AH. Bias in reporting clinical trials. *Br J Clin Pharmacol*. Aug 1998;46(2):147-150.

4. Simes RJ. Publication bias: the case for an international registry of clinical trials. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. Oct 1986;4(10):1529-1541.

5. Tam VC, Tannock IF, Massey C, Rauw J, Krzyzanowska MK. Compendium of unpublished phase III trials in oncology: characteristics and impact on clinical practice. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. Aug 10 2011;29(23):3133-3139.

6. Blumle A, Antes G, Schumacher M, Just H, von Elm E. Clinical research projects at a German medical faculty: follow-up from ethical approval to publication and citation by others. *J Med Ethics*. Sep 2008;34(9):e20.

7. de Jong JP, Ter Riet G, Willems DL. Two prognostic indicators of the publication rate of clinical studies were available during ethical review. *Journal of Clinical Epidemiology*. Dec 2010;63(12):1342-1350.

8. Decullier E, Chan AW, Chapuis F. Inadequate dissemination of phase I trials: a retrospective cohort study. *PLoS Med*. Feb 17 2009;6(2):e1000034.

9. Decullier E, Lheritier V, Chapuis F. Fate of biomedical research protocols and publication bias in France: retrospective cohort study. *BMJ*. Jul 2 2005;331(7507):19.

10. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA*. Jan 15 1992;267(3):374-378.

11. Easterbrook PJ, Matthews DR. Fate of research studies. *J R Soc Med*. Feb 1992;85(2):71-76.

12. Kasenda B, Schandelmaier S, Sun X, et al. Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications. *BMJ*. 2014;349:g4539.

13. Pich J, Carne X, Arnaiz JA, Gomez B, Trilla A, Rodes J. Role of a research ethics committee in follow-up and publication of results. *Lancet*. Mar 22 2003;361(9362):1015-1016.

14. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ*. Sep 13 1997;315(7109):640-645.

15. Sune P, Sune JM, Montoro JB. Positive outcomes influence the rate and time to publication, but not the impact factor of publications of clinical trial results. *PLoS One*. 2013;8(1):e54583.

16. von Elm E, Rollin A, Blumle A, Huwiler K, Witschi M, Egger M. Publication and non-publication of clinical trials: longitudinal study of applications submitted to a research ethics committee. *Swiss Med Wkly*. Apr 5 2008;138(13-14):197-203.

17. Massey PR, Wang R, Prasad V, Bates SE, Fojo T. Assessing the Eventual Publication of Clinical Trial Abstracts Submitted to a Large Annual Oncology Meeting. *Oncologist*. Mar 2016;21(3):261-268.

18. Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PloS one*. 2013;8(7):e66844.

19. ICMJE. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. Update December 2015. http://www.icmje.org/icmje-recommendations.pdf Last accessed on 2 November 2016.

20. Moorthy VS, Karam G, Vannice KS, Kieny MP. Rationale for WHO's new position calling for prompt reporting and public disclosure of interventional clinical trial results. *PLoS Med.* Apr 2015;12(4):e1001819.

21. Food and Drug Administration Amendments Act of 2007, Title VIII - Clinical trial databases. Public Law 110–85. September 27, 2007.

22. Amendment of the Medical Research Involving Human Subjects Act with regard to the evaluation of the act and recovery of incomplete implementation of guideline no. 2001/20/EG. Dossier 31452. Enacted as per 1 July 2012 (full text in Dutch only). The Hague, the Netherlands: Tweede Kamer der Staten Generaal (Dutch national Parliament).

23. European Federation of Pharmaceutical Industries and Associations (EFPIA) Position - Transparency of Information on Clinical Trials included in the Proposed EU Database (Article 78 of the Proposal for a Regulation on Clinical Trials). http://www.efpia.eu/uploads/Modules/Documents/efpia_key_messages_transparency_final_10.12.2012-updated_0.pdf. Last accessed on 2 November 2016.

24. Jones CW, Handler L, Crowell KE, Keil LG, Weaver MA, Platts-Mills TF. Non-publication of large randomized clinical trials: cross sectional analysis. *BMJ.* 2013;347:f6104.

25. Lampert A, Hoffmann GF, Ries M. Ten Years after the International Committee of Medical Journal Editors' Clinical Trial Registration Initiative, One Quarter of Phase 3 Pediatric Epilepsy Clinical Trials Still Remain Unpublished: A Cross Sectional Analysis. *PloS one.* 2016;11(1):e0144973.

26. Ramsey S, Scoggins J. Commentary: practicing on the tip of an information iceberg? Evidence of underpublication of registered clinical trials in oncology. *Oncologist.* Sep 2008;13(9):925-929.

27. Shamliyan T, Kane RL. Clinical research involving children: registration, completeness, and publication. *Pediatrics.* May 2012;129(5):e1291-1300.

28. Song F, Parekh-Bhurke S, Hooper L, et al. Extent of publication bias in different categories of research cohorts: a meta-analysis of empirical studies. *BMC Med Res Methodol.* 2009;9:79.

29. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet.* Apr 13 1991;337(8746):867-872.

30. van den Bogert CA, Souverein PC, Brekelmans CT, et al. Occurrence and determinants of selective reporting of clinical drug trials: design of an inception cohort study. *BMJ Open.* 2015;5(7):e007827.

31. ToetsingOnline. www.toetsingonline.nl. Last accessed on 5 August 2016.

32. http://www.ccmo.nl/en/accredited-mrecs Last accessed on 5 August 2016.

33. van den Bogert CA, Souverein PC, Brekelmans CT, et al. Non-Publication Is Common among Phase 1, Single-Center, Not Prospectively Registered, or Early Terminated Clinical Drug Trials. *PloS one.* 2016;11(12):e0167709.

34. Food and Drug Administration. Code of Federal Regulations 21CFR312.21. Phases of an investigation. Title 21, Volume 5, Revised as of April 1, 2015.

35. Nurgat ZA, Craig W, Campbell NC, Bissett JD, Cassidy J, Nicolson MC. Patient motivations surrounding participation in phase I and phase II clinical trials of cancer chemotherapy. *Br J Cancer.* Mar 28 2005;92(6):1001-1005.

2

36. Hole OP, Nitter-Hauge S, Cederkvist HR, Winther FO. An analysis of the clinical development of drugs in Norway for the year 2000: the completion of research and publication of results. *Eur J Clin Pharmacol.* Mar 2009;65(3):315-318.

37. Mattila T, Stoyanova V, Elferink A, Gispen-de Wied C, de Boer A, Wohlfarth T. Insomnia medication: do published studies reflect the complete picture of efficacy and safety? *Eur Neuropsychopharmacol.* Jul 2011;21(7):500-507.

38. Kasenda B, von Elm E, You J, et al. Prevalence, characteristics, and publication of discontinued randomized trials. *JAMA.* Mar 12 2014;311(10):1045-1051.

39. Moher D, Glasziou P, Chalmers I, et al. Increasing value and reducing waste in biomedical research: who's listening? *Lancet.* Sep 25 2015.

40. AllTrials Campaign. http://www.alltrials.net/. Accessed 4 August 2016

41. Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC. European Commission. Offic J Eur Union 2014;158:1-76.

42. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA.* May 26 2004;291(20):2457-2465.

43. Hemminki E. Study of information submitted by drug companies to licensing authorities. *Br Med J.* Mar 22 1980;280(6217):833-836.

44. Hewitt CE, Mitchell N, Torgerson DJ. Listen to the data when results are not significant. *BMJ.* Jan 5 2008;336(7634):23-25.

45. Stewart LA, Parmar MK. Bias in the analysis and reporting of randomized controlled trials. *Int J Technol Assess Health Care.* Spring 1996;12(2):264-275.

46. Law ignored, patients at risk. December 13, 2015. http://www.statnews.com/2015/12/13/clinical-trials-investigation/. Last accessed on 29 March 2016.

47. Chen R, Desai NR, Ross JS, et al. Publication and reporting of clinical trial results: cross sectional analysis across academic medical centers. *BMJ.* 2016;352:i637.

48. Van den Bogert CA, Cohen AF. Need for a proactive and structured approach to risk analysis when designing phase I trials. *BMJ.* 2015;351:h3899.

49. Royal Statistical Society statement on publication of the study-protocol BIA-102474-101 for the French "first-in-man" trial in healthy volunteers. http://www.rss.org.uk/Images/PDF/about/press-releases/2016-01-22-rss-statment-BIA-102474-101-french-first-trial-in-healthy-volunteers.pdf, 22 January 2016. Last accessed on 29 March 2016.

50. Salzberg M. First-in-Human Phase 1 Studies in Oncology: The New Challenge for Investigative Sites. *Rambam Maimonides Med J.* Apr 2012;3(2):e0007.

51. Ioannidis JP, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet.* Jan 11 2014;383(9912):166-175.

52. Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Educ Prev.* Feb 1997;9(1 Suppl):15-21.

53. Huser V, Cimino JJ. Evaluating adherence to the International Committee of Medical Journal Editors' policy of mandatory, timely clinical trial registration. *J Am Med Inform Assoc.* Jun 2013;20(e1):e169-174.

Chapter 2.4

Discrepancies between protocols and publications of clinical drug trials

Cornelis A van den Bogert, Patrick C Souverein, Cecile TM Brekelmans, Susan WJ Janssen, Gerard H Koëter, Hubert GM Leufkens, Lex M Bouter

ABSTRACT

Objectives
To identify the occurrence and determinants of protocol-publication discrepancies in clinical drug trials.

Methods
All published clinical drug trials reviewed by the Dutch Institutional Review Boards in 2007 were analyzed. Discrepancies between trial protocols and publications were measured among key reporting aspects. We evaluated the association of trial characteristics with discrepancies in primary endpoints by calculating the risk ratio (RR) and 95% confidence interval (CI).

Results
Of the 334 published trials, 32 (9.6%) had a protocol/publication discrepancy in the primary endpoints. Among the subgroup of randomized controlled trials (RCTs; N=204), 12 (5.9%) had a discrepancy in the primary endpoint. Investigator-initiated trials with and without industry (co-) funding were associated with having discrepancies in the primary endpoints compared to industry-sponsored trials (RR 3.7; 95% CI 1.4-9.9 and RR 4.4; 95% CI 2.0-9.5, respectively). Furthermore, other than phase 1-4 trials (vs. phase 1; RR 4.6; 95% CI 1.1-19.3), multicenter trials also conducted outside the EU (vs. single center; RR 0.2; 95% CI 0.1-0.6), not prospectively registered trials (RR 3.3; 95% CI 1.5-7.5), non-RCTs (vs. superiority RCT; RR 2.4; 95% CI 1.2-4.8) and, among the RCTs, crossover compared to a parallel group design (RR 3.7; 95% CI 1.1-12.3) were significantly associated with having discrepancies in the primary endpoints.

Conclusions
Improvement in completeness of reporting is still needed, especially among investigator-initiated trials and non-RCTs. To eliminate undisclosed discrepancies, trial protocols should be available in the public domain at the same time when the trial is published.

2

INTRODUCTION

Selective reporting is considered to be the most important cause of the poor reproducibility of biomedical research[1]. If mainly the positive results of a study are published, this may lead to overrepresentation of positive results and conclusions in the scientific literature[2]. Ignoring negative results can cause research waste, as futile experiments may be unnecessarily repeated [3]. Moreover, an inadequate description of the protocol of a study can frustrate replication of the study[4,5]. Transparency of the process from study protocols until publication remains therefore paramount in the responsible conduct of research. Complete and unbiased publication of clinical trials is an ethical and scientific obligation as recommendations and conclusions derived from clinical trials are often translated into clinical guidelines, and human participants were involved in obtaining the results[6].

One type of selective reporting is non-publication[7]. Evidence across medical and geographical areas shows that approximately half of the clinical trials that are conducted, are not being published in the scientific literature [8]. Another type is selective publication, meaning that at least some results are published, but with undisclosed discrepancies between the trial protocol and the publication. The first empirical study investigating the problem of selective publication found that 62% of the trials had at least one primary endpoint that was discrepant between the trial protocol and the trial publication[9]. As the main conclusions and recommendations of trials will be based on their primary outcome, this finding suggests that a substantial proportion of clinical evidence is biased due to selective publication. Other studies also showed an alarming amount of selective publication regarding subgroup analyses, sample size calculations and sponsorship acknowledgement[10-13].

While the existence of selective publication has been convincingly established among clinical trials starting 15-20 years ago[9,10], its occurrence may have decreased due to subsequent countermeasures. Governments, journals, pharmaceutical companies, and research communities have implemented requirements for trial registration and data sharing[14-20]. However, more recent evidence suggests that only limited progress has been made[10]. Empirical evidence is very limited on whether other aspects of trials are transparently reported, such as the selection criteria, sample size, subgroup or other additional analyses, and the methods used for data analysis. Therefore, we studied the occurrence of protocol-publication discrepancies, determinants of discrepancies in primary endpoints, and the association between discrepancies and the direction of trial conclusions in a cohort of clinical drug trials.

METHODS

The design of the study has been published before[7]. In short, we selected all clinical drug trials that were reviewed by the Dutch accredited Institutional Review Boards in 2007 and we followed these trials until publication as peer-reviewed article in the scientific literature. The results of the study on non-publication have been published and showed that of the 574 trials in the cohort, 240 (42%) remained unpublished[21]. For this follow-up study, we included the 334 trials in the cohort of which we found at least one publication by January 2016 in the scientific literature presenting results (figure 2.4.1). The data source was ToetsingOnline, a database containing all clinical drug trials submitted to accredited Institutional Review Boards (IRBs) in the Netherlands, overseen by the competent authority (the Central Committee on Research Involving Human Subjects, CCMO). Hence, the cohort consists of all clinical drug trials that were IRB-reviewed in the Netherlands in 2007. The data sources for the discrepancy assessment were the IRB-files of the CCMO including the original trial protocols and substantial amendments (as required by law, these documents are submitted to both the IRB and the CCMO before start of the trial or before implementation of the amendment). We searched Pubmed, Embase and Google Scholar for publications of the trials in scientific journals. All publications containing results of the trials were downloaded as full-text. The publication search was conducted in January and February 2016. Thus, the minimal follow-up between IRB-approval and publication was 8 years (December 2007 – January 2016).

We identified discrepancies between the most recently IRB-approved trial protocol (including IRB-approved substantial amendments) and the publications reporting results of the trial at issue. If multiple publications of the same trial were identified, all publications were included in the assessment. Discrepancies were scored by comparing the full-text of the original protocol with the full-texts of all identified publications of the trial. Five categories of protocol-publication discrepancies were measured: endpoints (the operationalization of events, symptoms, biomarkers etc. that were measured in the trial); trial objectives (the general conceptual goals of the trial as stated in the introduction of the protocols and the publications); selection criteria; sample size; and sponsor acknowledgement. Two additional protocol-publication discrepancies were only scored among the randomized controlled trials (RCTs) in the cohort: discrepancies in additional or subgroup analyses, and discrepancies in the method used for the data analysis. The method used for data analysis was defined as how the trial arms were compared. For example, using the intention-to-treat, or the per protocol approach for the analysis (and whether the definition of the analysis population was similar

in the protocol and in the publication). More details on the protocol-publication categories are provided in the supplementary methods and results, table S2.4.1.
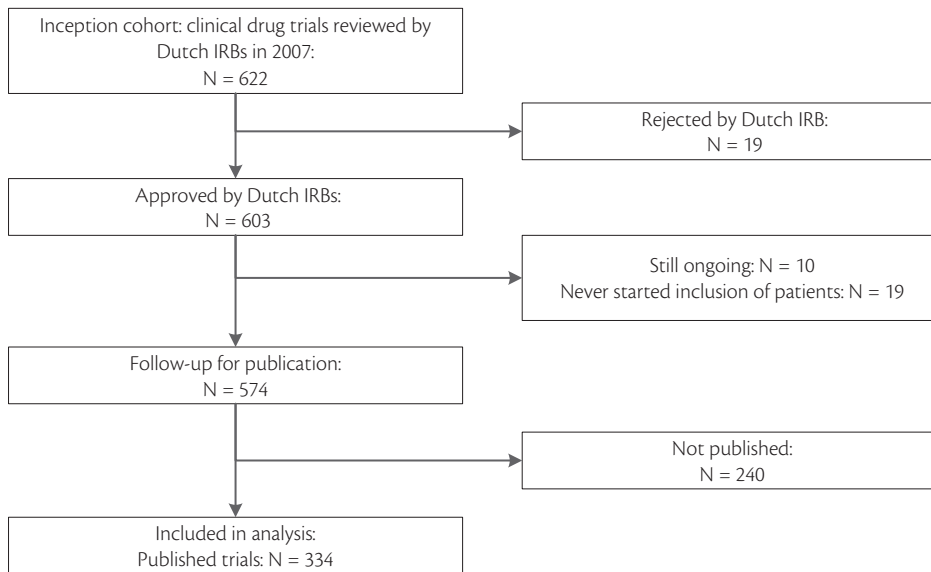

FIGURE 2.4.1 Trials included in the analysis.

Disclosure was the leading principle in scoring the discrepancies. Aspects were only scored as being discrepant if no reason for the discrepancy was provided in the publication, and we could not find another reason for the discrepancy that was disclosed to the IRB (such as approved substantial amendments). For example, if a primary endpoint was added in a publication compared to the trial protocol, and the publication also explained this addition, it was not considered discrepant. Similarly, a post hoc subgroup analysis was only considered discrepant if the post hoc nature of the subgroup analysis was not stated in the publication. In addition, if the trial was discontinued before the planned end of follow-up, and this was reported in the publication, we did not score the lower sample size as discrepancy. If multiple methods for data analysis were described in the publication and only one was specified in the protocol, we only scored this as being a discrepancy if the publication did not state which method was specified in the protocol. In case that multiple publications were found of one trial, omissions were only scored as discrepancy if the omitted item was not reported in any of the publications of the trial. Additions were scored as discrepancy if not labeled as post hoc in all publications of the trial. If we found in one of the publications an unexplained change compared to the protocol, it was scored as a discrepancy, regardless whether it was reported correctly in the other publications of the trial.

Trial characteristics were extracted from the ToetsingOnline database, from the form that all trial applicants filled out at the time of submission of the trial application for IRB-review. This form is mandatory and identical for all IRBs in the Netherlands. Other trial characteristics were prospective registration in the international registries of clinicaltrials.gov or ISRCTN, and whether the trial was completed as planned or discontinued. Other trial characteristics were the trial design (RCT superiority, RCT non-inferiority, or non-RCT/exploratory pharmacology), and, only among the RCTs, the treatment arms (parallel group or crossover). Associations between these trial characteristics and protocol-publication discrepancies in the primary endpoints were evaluated.

Among the subgroup of RCTs, we categorized the direction of conclusion of the trials as positive or negative, as formulated in the publications. The direction of conclusion was positive if the trial results supported the trial objectives or hypotheses as stated in the protocol (for example, drug X is superior compared to placebo against disease Y). If the conclusion section of the publication stated that the results were negative, non-significant, or inconclusive, the direction of conclusion was classified as negative. If more than one publication was found of a trial, the first publication of the completed trial that reported primary endpoints was used to classify the direction of conclusion.

The protocol-publication discrepancies were described using univariate analysis, stratified for RCTs and non-RCTs. Randomized trials with exploratory objectives (for example, phase 1 trials investigating pharmacology, safety and tolerability) were included in the non-RCT stratum. For the discrepancies in endpoints, objectives and additional/subgroup analyses, we merged the outcome variable by calculating the sum total of primary endpoint discrepancies and primary objective discrepancies, and the sum total of discrepancies in additional or subgroup analyses.

We analyzed the association between the trial characteristics that were considered as being potential determinants of protocol-publication discrepancies. In addition to the sponsor type, we analyzed the trial characteristics that were significantly associated with non-publication in the same cohort[21]: phase, centers involved, prospective registration, and completion. Furthermore, in line with previous studies[9,10], we also analyzed the association of the trial design and the treatment arms with discrepancies in the primary endpoints. The protocol of our study[7] prescribed the analysis of determinants for all protocol-publication discrepancies separately. In this paper, we focus on determinants of the discrepancies in the primary endpoints, which are most likely to influence the direction of conclusions of the trials. Then, we analyzed the association between the protocol-publication discrepancies and the direction of conclusions of the trials.

To estimate the overall associations, we used Pearson's chi-square test and indicate the associations of p <0.01 and p < 0.05. Furthermore, risk ratios (RRs) and their 95% confidence interval (CIs) were calculated to estimate the direction and precision of the associations. If zero outcomes (or zero reciprocal-outcomes) were observed in categories with low numbers of trials, the RR was not calculated as a zero cell count will strongly bias the association towards statistical significance. The protocol of our study prescribed also multivariable logistic regression analysis. However, due to the relatively low number of discrepancies in the primary endpoints (32), the precision of the regression coefficients would have been low. Therefore, we decided to omit multivariable analysis.

One investigator (CAB) performed the discrepancy scoring for all trials. A second investigator (PCS) examined the reliability of the discrepancy scoring method. The protocol prescribed an additional double-check of 10% of the cohort and subsequently a randomly selected 20 trials. After comparing the seven discrepancy categories of the initial 35 trials selected for crosschecking (245 data entries in total), three data entries were changed after discussion. These included one discrepancy in the selection criteria, one discrepancy in the secondary endpoint, and one discrepancy in the subgroup or additional analyses. Thus, the inter-rater agreement was (1-(3/245))*100 = 99%, with no disagreements about discrepancies in the primary endpoints. Based on the inter-rater agreement of 99%, we concluded there was sufficient proof of reliability of the scoring procedure and that the double-check could be restricted to the randomly selected 35 trials (10%). We included this protocol deviation in table S2.4.2 of the supplementary methods and results.

RESULTS

Of the 334 trials that were published by January 2016 (figure 2.4.1), we identified 506 articles. Of 91% of these trials, we found one or two articles (supplementary methods and results, table S2.4.3). The characteristics of the 334 trials are summarized in table 2.4.1. The trials were mostly industry-sponsored (62.3%), phase 3 (37.4%) and/ or international multicenter (16.8% was also conducted in other EU countries, and 40.7% was also conducted outside the EU). Oncology was the largest disease area (22.5%), and most trials (51.8%) were not prospectively registered at clinicaltrials. gov or ISRCTN. A small proportion (10.2%) was discontinued before the planned end of recruitment and/or follow-up. Sixty-one percent were RCTs (50.6% superiority and 10.5% non-inferiority), and most trials had a parallel group design. Almost half of the trials (47.9%) planned to include less than 100 participants.

TABLE 2.4.1 Characteristics of the analyzed trials.

	Number of trials in analysis	
Total clinical trials in cohort	334 (100%)	% (of 334)
Characteristics		
Sponsor		
Pharmaceutical industry	208	62.3%
Investigator (industry (co-)funded)	37	11.1%
Investigator (no industry funding involved)	89	26.6%
Phase		
Phase 1	41	12.3%
Phase 2	78	23.4%
Phase 3	125	37.4%
Phase 4	32	9.6%
Other than phase 1-4*	58	17.4%
Centers		
Single center	113	33.8%
Multi center only in the Netherlands	29	8.7%
Multi center in the Netherlands and the EU	56	16.8%
Multi center in the Netherlands and outside the EU	136	40.7%
Disease area		
Oncology	75	22.5%
Endocrine diseases	40	12.0%
Neurological diseases (including analgesia and anesthesia trials)	36	10.8%
Infectious diseases (including vaccine trials)	32	9.6%
Cardiovascular diseases	29	8.7%
Respiratory diseases	25	7.5%
Other disease areas	22	6.6%
Musculoskeletal diseases	19	5.7%
Mental and behavioral disorders	17	5.1%
Hematological and immunological diseases	17	5.1%
Digestive system diseases	12	3.6%
Genitourinary system diseases	10	3.0%
Prospective registration[‡]		
Prospectively registered	161	48.2%
Not (prospectively) registered	173	51.8%
Completion		
Completed as planned	300	89.8%
Terminated early	34	10.2%

TABLE 2.4.1 Characteristics of the analyzed trials. (continued)

	Number of trials in analysis	
Design[†]		
RCT; superiority	169	50.6%
RCT; non-inferiority	35	10.5%
Non-RCT and/or exploratory pharmacology trial	130	38.9%
Treatment arms		
Parallel group	226	67.7%
Cross-over	39	11.7%
Single arm	69	20.7%
Number of participants planned in the trial protocol		
<100	160	47.9%
≥100-500	89	26.6%
≥500-1000	50	15.0%
≥1000	35	10.5%

Table 2.4.2 shows an overview of the protocol-publication discrepancies that were measured in all 334 trials. Omissions (N = 17; 5.1%) and changes (N = 14; 4.2%) of the primary endpoint were more common than additions (N = 1; 0.3%). The most common discrepancies were in secondary endpoints: 89 (43.6%) of the RCTs and 48 (36.9%) of the non-RCTs had no discrepancy in the primary endpoints but a discrepancy in the secondary endpoints. Discrepancies in the sample size were mainly due to inclusion of <80% of the sample as calculated in the protocol, which occurred in 10 (4.9%) of the RCTs and in 22 (16.9%) of the non-RCTs). Three (1.5%) of the RCTs and three (2.3%) of the non-RCTs included more than 120% of the sample size as calculated in the protocol.

Table 2.4.3 summarizes the discrepancy categories that were only measured among the subgroup of 204 RCTs. Among the 204 RCTs, 91 (44.6%) had a discrepancy in the subgroup analysis. Furthermore, 21 (10.3%) of the RCTs had a discrepancy in the methods used for data analysis. None of the trials had a discrepancy in sponsorship acknowledgements. Seventy-eight trials (23.3%) had no discrepancy at all (table S2.4.4, supplementary methods and results). In 36 (17.6%) of the RCT-protocols and/or publications, the methods used for data analysis were not specified (table 2.4.3). In 21 of the 204 RCTs (10.3%), it was only missing in the protocol, in nine RCTs (4.4%) only in the publications, and in six (2.9%) both in the protocol and in the publications. The information on the methods used for data analysis was missing in 30 (55.6%) of the protocols and/or publications of the 54 investigator-initiated RCTs, and in 6 (4.0%) of the 150 industry-sponsored RCTs. The other discrepancy categories were missing in the protocols and/or publications of zero to four trials (0-1.5%).

TABLE 2.4.2 Occurrence of protocol-publication discrepancies stratified for trial design.

	All trials	RCTs	Non-RCTs[‡]
	N (%)	N (%)	N (%)
Number of trials assessed	**334 (100%)**	**204 (100%)**	**130 (100%)**
Discrepancies in endpoints			
Primary endpoint added in publication	1 (0.3%)	0 (0.0%)	1 (0.8%)
Primary endpoint omitted in publication	17 (5.1%)	3 (1.5%)	14 (10.8%)
Primary endpoints changed in publication	14 (4.2%)	9 (4.4%)	5 (3.8%)
Sum total of discrepancies in primary endpoint*	32 (9.6%)	12 (5.9%)	20 (15.4%)
No discrepancy in primary endpoint, but discrepancy in secondary endpoint	137 (41.0%)	89 (43.6%)	48 (36.9%)
No discrepancies in endpoints	163 (48.8%)	102 (50.0%)	61 (46.9%)
No information in protocol and/or publication on endpoints	2 (0.6%)	1 (0.5%)	1 (0.8%)
Discrepancies in objectives			
Primary objective added in publication	2 (0.6%)	2 (1.0%)	0 (0.0%)
Primary objective omitted in publication	12 (3.6%)	2 (1.0%)	10 (7.7%)
Primary objective changed in publication	11 (3.3%)	6 (2.9%)	5 (3.8%)
Sum total of discrepancies in primary objective*	25 (7.5%)	10 (4.9%)	15 (11.5%)
No discrepancy in primary objective, but discrepancy in secondary objective	64 (19.2%)	47 (23.0%)	17 (13.1%)
No discrepancies in objectives	241 (72.2%)	144 (70.6%)	97 (74.6%)
No information in protocol and/or publication on objectives	4 (1.2%)	3 (1.5%)	1 (0.8%)
Discrepancies in selection criteria			
Changed in publication	37 (11.1%)	21 (10.3%)	16 (12.3%)
No discrepancies in selection criteria	295 (88.3%)	183 (89.7%)	112 (86.2%)
No information in protocol and /or publication on selection criteria	2 (0.6%)	0 (0.0%)	2 (1.5%)
Discrepancies in sample size			
< 80% of sample size as calculated in protocol included	32 (9.6%)	10 (4.9%)	22 (16.9%)
> 120% of sample size as calculated in protocol included	6 (1.8%)	3 (1.5%)	3 (2.3%)
No discrepancies in selection criteria	296 (88.6%)	191 (93.6%)	105 (80.8%)
No information in protocol and/or publication on sample size	0 (0.0%)	0 (0.0%)	0 (0.0%)

RCT = randomized controlled trial.

* Sum total of all discrepancies in primary endpoints or primary objectives

[‡] Randomized exploratory pharmacology trials were also included in the non-RCT stratum.

TABLE 2.4.3 Occurrence of protocol-publication discrepancies that were only scored among the subgroup of randomized controlled trials (RCTs).

	N (%)
Number of trials assessed	**204 (100%)**
Discrepancies in additional or subgroup analyses	
Additional/subgroup analysis added in publication	25 (12.3%)
Additional/subgroup analysis omitted in publication	62 (30.4%)
Additional/subgroup analysis changed in publication	4 (2.0%)
Sum total of discrepancies in additional or subgroup analysis*	91 (44.6%)
No discrepancies in additional or subgroup analysis	111 (54.4%)
No information in protocol and/or publication on subgroup analysis	2 (1.0%)
Discrepancies in method used for data analysis	
Method used for data analysis changed in the publication	21 (10.3%)
No discrepancies in method used for data analysis	147 (72.1%)
No information in protocol and/or publication on data analysis	36 (17.6%)

* Sum total of all discrepancies in additional or subgroup analysis

Associations between trial characteristics and discrepancies in the primary endpoints are shown in table 2.4.4. Investigator-initiated trials were associated with a higher likelihood to have a discrepancy in the primary endpoints compared to industry-sponsored trials, whether or not the industry was involved as (one of) the funder(s) of the trial (RR 3.7, 95% CI 1.4-9.9 and RR 4.4, 95% CI 2.0-9.5, respectively). Furthermore, trials not being classified as phase 1-4 had a higher likelihood of discrepancies in the primary endpoints compared to phase 1 trials (RR 4.6, 95% CI 1.1-19.3). Multicenter trials also conducted outside the EU had a lower likelihood of having a discrepancy in the primary endpoints compared to single center trials (RR 0.2, 95% CI 0.1-0.6). Trials that were not prospectively registered in clinicaltrials.gov or the ISRCTN registry were more likely to have a discrepancy in the primary endpoints compared to trials that were prospectively registered (RR 3.3, 95% CI 1.5-7.5). Compared to superiority RCTs, non-RCTs had a higher likelihood to have a discrepancy in the primary endpoint (RR 2.4, 95% CI 1.2-4.8). This association was not observed when comparing superiority RCTs to non-inferiority RCTs (RR 0.4, 95% CI 0.1-3.3). Finally, crossover RCTs had a higher likelihood of discrepancies in the primary endpoint compared to parallel group RCTs (RR 3.7, 95% CI 1.1-12.3).

TABLE 2.4.4 Association between trial characteristics and protocol-publication discrepancies in primary endpoints.

	N (%)	X^2	RR (95% CI)
Total clinical trials in analysis	334 (100%)		
Discrepancies in primary endpoints[‡]	32 (9.6%)		
Sponsor		17.82*	
Pharmaceutical industry (n = 208)	9 (4.3%)		ref
Investigator (industry (co-)funded) (n = 37)	6 (16.2%)		3.7 (1.4-9.9)
Investigator (no industry funding involved) (n = 89)	17 (19.1%)		4.4 (2.0-9.5)
Phase		17.04*	
Phase 1 (n = 41)	2 (4.9%)		ref
Phase 2 (n = 78)	6 (7.7%)		1.6 (0.3-7.5)
Phase 3 (n = 125)	6 (4.8%)		1.0 (0.2-4.7)
Phase 4 (n = 32)	5 (15.6%)		3.2 (0.7-15.4)
Other than phase 1-4[§] (n = 58)	13 (22.4%)		4.6 (1.1-19.3)
Centers		15.42*	
Single center (n = 113)	19 (16.8%)		ref
Multi center only in the Netherlands (n = 29)	5 (17.2%)		1.0 (0.4-2.5)
Multi center in the Netherlands and the EU (n = 56)	3 (5.4%)		0.3 (0.1-1.0)
Multi center in the Netherlands and outside the EU (n = 136)	5 (3.7%)		0.2 (0.1-0.6)
Prospective registration[‖]		9.83*	
Prospectively registered (n = 161)	7 (4.3%)		ref
Not (prospectively) registered (n = 173)	25 (14.5%)		3.3 (1.5-7.5)
Completion		0.03	
Completed as planned (n = 300)	29 (9.7%)		ref
Discontinued before planned end (n = 34)	3 (8.8%)		0.9 (0.3-2.8)
Design[#]		8.72[†]	
RCT; superiority (n = 169)	11 (6.5%)		ref
RCT; non-inferiority (n = 35)	1 (2.9%)		0.4 (0.1-3.3)
Non-RCT and/or exploratory pharmacology trial (n = 130)	20 (15.4%)		2.4 (1.2-4.8)
Subgroup of RCTs	204 (100%)		
Discrepancies in primary endpoints[‡]	12 (5.9%)		
Treatment arms		4.64[†]	
Parallel group (n = 187)	9 (4.8%)		ref
Crossover (n = 17)	3 (17.6%)		3.7 (1.1-12.3)

RCT = randomized controlled trial; RR = risk ratio; CI = confidence interval. * P-value < 0.01 (based on Pearson's chi-square) [†] P-value between 0.01 and 0.05 (based on Pearson's chi-square). [‡] The sum total of discrepancies in primary endpoints (N = 32; see table 2), or for the subgroup of RCTs (N=12). [§] Trials carried out using medicinal products in connection with objectives other than those referred to in the phase definitions 1-4. Such trials are not intended primarily to provide information about the product itself, but a medicinal product is needed in order to address the objective of the trial [31]. [‖] Prospective registration was defined as registration of the trial at the international public registers www.clinicaltrials.gov or www.isrctn.com, at latest one month after IRB-approval. [#] Exploratory pharmacology trials that involved randomization, but no formal hypothesis testing (which is common in, for example, phase 1 randomized dose-escalation trials) were also excluded from the RCT-subgroup.

Table 2.4.5 shows the association between the protocol-publication discrepancies and a positive direction of the trial conclusions. In none of the discrepancy categories, having a protocol-publication discrepancy in that category was associated with a positive direction of trial conclusions.

DISCUSSION

We found that 9.6% of all clinical drug trials, and 5.9% of the RCTs, in our study had a protocol-publication discrepancy in the primary endpoints. This is a substantially lower proportion than reported by the two previous studies investigating this issue. Chan et al. found discrepancies in primary endpoints among 62% of RCTs[9]. Berendt et al. conducted a study among academic (investigator-initiated) trials and found discrepancies in primary endpoints in 38% and 43% in non-RCTs and RCTs, respectively[10]. In the subgroup of the 126 academic trials in our cohort, discrepancies in primary endpoints were found in 10 out of 54 RCTs (19%), and in 13 out of 72 non-RCTs (18%). This finding suggests that also the reporting of academic trials has been improved. Furthermore, both in RCTs and in non-RCTs, protocol-publication discrepancies were substantial in secondary endpoints: 89 (44%) and 48 (37%), respectively. These proportions were also considerably lower than those of the recent COMPare initiative[22], which reported discrepancies in endpoints among 87% of the trials (not differentiating between primary and secondary endpoints). Discrepancies in primary objectives were found in 7% (N = 25) of the trials, and in 5% (N = 10) of the RCT subgroup. This is also lower than the previous study that investigated this discrepancy[10]. In line with secondary endpoints, discrepancies in secondary objectives occurred in 19% (N = 64) of the trials and in 23% (N = 47) of the RCTs. We found a discrepancy in selection criteria in 11% (N = 37) of the trials, and in 10% (N = 21) of the RCTs. To our knowledge, our study is the first to investigate this.

TABLE 2.4.5 Association between protocol-publication discrepancies of the subgroup of randomized controlled trials (RCTs) and the direction of conclusion.

	RCTs with positive direction of conclusion N (%)	X^{2*} RR (95% CI)
Total RCTs	204 (100%)	
Total number of RCTs with positive direction of conclusion	141 (69.1%)	
Endpoints		4.29
No discrepancies (n = 102)	75 (73.5%)	ref
Primary endpoint added in publication (n = 0)	-	-
Primary endpoint omitted in publication (n = 3)	2 (66.7%)	0.9 (0.4-2.0)
Primary endpoint changed in publication (n = 9)	4 (44.4%)	0.6 (0.3-1.3)
Secondary endpoint omitted/added/changed in publication (n = 89)	59 (66.3%)	0.9 (0.7-1.1)
No information in protocol (n = 1)	1 (100%)	-[†]
Objectives		8.37
No discrepancies (n = 144)	105 (72.9%)	ref
Primary objective added in publication (n = 2)	2 (100%)	-[†]
Primary objective omitted in publication (n = 2)	0 (0%)	-[†]
Primary objective changed in publication (n = 6)	4 (66.7%)	0.9 (0.5-1.6)
Secondary objective omitted/added/changed in publication (n = 47)	28 (59.6%)	0.8 (0.6-1.1)
No information in protocol (n = 3)	2 (66.7%)	0.9 (0.4-2.0)
Selection criteria		0.06
No discrepancies (n = 183)	126 (68.9%)	ref
Changed in publication (n = 21)	15 (71.4%)	1.0 (0.8-1.4)
No information in protocol (n = 0)	-	-

Discrepancies in the sample size were also found in 11% (N = 38) of the trials, and in 6% (N = 13) of the RCTs). This finding is considerable lower than reported in a previous study, which found discrepancies in sample size calculations in 53% of the RCTs[23].

Among 91 (45%) of the 204 RCTs, we also found discrepancies in subgroup or other additional analyses. In this discrepancy category, omissions of subgroup or additional analysis that were planned in the protocol were most common. Omissions of planned subgroup analyses are, to our knowledge, not investigated in previous studies. The 25 RCTs (12%) that added an unplanned subgroup analysis and did not label it as being post hoc is lower than a previous study that found 35%[11]. Finally, 21 (10%) of the RCTs changed the method used for data analysis. This is also considerably lower compared to findings by a previous study[23].

TABLE 2.4.5 Association between protocol-publication discrepancies of the subgroup of randomized controlled trials (RCTs) and the direction of conclusion. (continued)

	RCTs with positive direction of conclusion N (%)	X^{2*} RR (95% CI)
Sample size		3.26
No discrepancies (n = 191)	133 (69.6%)	ref
< 80% of sample size as calculated in protocol included (n = 10)	6 (60.0%)	0.9 (0.5-1.4)
>120% of sample size as calculated in protocol included (n = 3)	2 (66.7%)	1.0 (0.4-2.1)
No information in protocol (n = 0)	-	-
Additional/subgroup analyses		3.06
No discrepancies (n = 111)	76 (68.5%)	ref
Additional/subgroup analysis added in publication (n = 25)	18 (72.0%)	1.1 (0.8-1.4)
Additional/subgroup analysis omitted in publication (n = 62)	41 (66.1%)	1.0 (0.8-1.2)
Additional/subgroup analysis changed in publication (n = 4)	4 (100%)	-[†]
No information in protocol (n = 2)	2 (100%)	-[†]
Methods used for data analysis		1.85
No discrepancies (n = 147)	101 (68.7%)	ref
Changed in publication (n = 21)	17 (81.0%)	1.2 (0.9-1.5)
No information in protocol (n = 36)	23 (63.9%)	0.9 (0.7-1.2)

[notes to table 2.4.5] RCT = randomized controlled trial; RR = risk ratio; CI = confidence interval. *None of the Pearson's chi-square tests indicated a statistically significant association (p < 0.05) between protocol-publication discrepancies and a positive direction of trial conclusions. [†] In case 100%, or 0%, of the RCTs within a category with a low number of trials had a positive direction of conclusion, the risk ratio was not calculated because a zero cell count would bias the estimation of the 95% CI towards statistical significance.

The reason for the lower occurrence of discrepancies compared to the three previous cohort studies[9-11] may be that more clinical investigators and journals are aware of the importance of complete and accurate reporting of all protocol aspects. In 2007, the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) Statement initiative was launched. Furthermore, the CONsolidated Standards of Reporting Trials (CONSORT) statement was updated in 2010. The broad attention to and implementation of these initiatives by the major medical journals might have contributed to better reporting[19,24-29]. Another reason could be related to the information that was available to explain the observed differences between protocols and publications. If a potential discrepancy was identified, we exhaustively searched substantial amendments and follow-up publications to explain the difference. If these explanations were available, for example in online supplementary files or in publications other than the main results publication, we did not consider it as being a discrepancy. Hence, it might have made a difference whether the research team was determined to find as many discrepancies

as possible, or find as many explanations for potential discrepancies as possible. Another reason for the difference with two of the previous cohorts is that we included all available publications in the assessment, whereas these studies only included one publication for each trial protocol[9,11,23]. The third previous cohort study[10] had only a follow-up from IRB-approval until publication of five years, which is likely too short to identify all relevant (additional) publications[30].

We replicated the previous findings that non-RCTs/exploratory trials have more discrepancies than RCTs[10], and that, among the RCTs, trials with a crossover treatment arm have more discrepancies compared to parallel group treatment arms[9]. Furthermore, the analysis of trial characteristics suggests that discrepancies in primary endpoints mainly occur in small, local, investigator-initiated trials not in the context of drug product development (other than phase 1-4[31]), that also were less likely to be prospectively registered. These characteristics often coincide. A reason for this may be that such trials are more flexible in the choice of endpoints, as the protocol and subsequent publications will not be reviewed by the drug marketing authorization authorities (who request standardized endpoints for a given disease area[32]). These analyses elucidate trial types that have remained outside the reach of the past initiatives and countermeasures against selective reporting.

No discrepancy categories were significantly associated with a positive direction of conclusion of the RCT subgroup. This suggests that the observed discrepancies have not been introduced in the publications to change the overall direction of conclusion of the trials. However, this was only assessed using a binary classification of direction of conclusion, leaving no room for nuances. In particular the discrepancies in primary and secondary endpoints may still have led to a 'more positive', or 'less negative' conclusion, thereby introducing reporting bias[22,33]. If investigators measure several endpoints, and can decide afterwards which to report, the likelihood is high that those endpoints are reported that fit expectations or desires[34]. If objectives are discrepant, the publication may fail to provide an accurate description of the original rationale and research question of the trial. This can be relevant for the interpretation of results. Discrepancies in the sample size can bias the interpretation of results, as the likelihood of erroneous chance findings is high if the sample size is too small[35]. Discrepancies in the methods used for data analysis can also be a way to spin the interpretation of results towards the preferred conclusion, for example by excluding or including outliers or cases with partially missing data entries. Not reporting planned subgroup analyses occur likely because of the absence of effect, and unplanned subgroup analyses were probably added post hoc because there was an (unexpected) effect. Although the latter can serendipitously lead to important discoveries, their exploratory

nature should be clearly acknowledged when they are reported[2,36]. Furthermore, protocol-publication discrepancies in the selection criteria can affect the ethical justification to include certain participants in the trial. For example, a protocol in the cohort prescribed inclusion of only patients with a given tumor characterized as grade 4 (the tumor grade indicating the most severe grade of illness). The publication reported inclusion of patients with tumor grade 3-4. However, as the protocol stated only the inclusion of grade 4 tumors, the IRB had approved the trial to be conducted specifically in the population with grade 4. The IRB had not considered grade 3 tumors in their evaluation, and investigators were therefore not permitted to include these patients. In addition, including more participants than needed according to the protocol can also be unethical. The research question could then have been answered at the cost of a smaller number of participants being exposed to risks and burdens of the trial[37].

In our cohort, the magnitude of non-publication likely exceeded the magnitude of selective publication in terms of causing research waste and publication bias. Of the 574 drug trials initially followed until publication, 42% remained unpublished[21], and of the published trials, 32 (10%) had discrepancies in the primary endpoints). Nevertheless, the identified discrepancies could still have introduced spin and bias in the trial publications[38]. Trial publications should therefore become more transparent and provide a clear track record of the process of a clinical trial, from the initial research protocol until the publications presenting the results[39]. Some journals published the trial protocol as well as protocol amendments as online supplement, but this was rather an exception than common practice. To further facilitate independent interpretation of protocol deviations, simple checklists can indicate which part of the protocol changed, when and why, and to which extent this may have influenced the conclusions[40]. This discrepancy checklist could then also be included in the assessment of bias (such as outcome reporting bias), which should be done when the trials are included in systematic reviews[4,41,42]. The finding that 20% of the RCTs missed information in the protocol and/or the publication about the methods used for data analysis incites IRBs and journals to always request this important information[20,24]. This attention is especially needed for the investigator-initiated RCTs.

A strength of our study is that the selection of protocols was not limited to those that are publicly available, thereby avoiding selection bias. We had access to all clinical drug trials that were submitted to an IRB in 2007. Further, we did not limit our discrepancy assessment to the endpoints, but assessed seven essential trial aspects that should be consistent between trial protocol and subsequent publications. As we included trials across all medical specialties, and 57% of the trials were multicenter international trials, our findings can be considered as

being generalizable across geographical and medical areas. Although a higher sample size would have enabled a more precise conclusion, the number of 334 trials included in our cohort is higher than most previous studies, with the exception of the study by Kasenda et al[9-11,23]. A limitation of our study was that we might have missed some documents explaining the discrepancies that were not included in the CCMO-archive. However, this missing information would then be incidental and therefore unlikely systematic or differential. And, if substantial amendments were missing in the archives, a record of these as well as non-substantial amendments should nevertheless have been provided in the publication, thus discrepancies scored as a result from these missing documents can be justified. Another limitation is that we might have missed some publications that were published after the follow-up period, or were missed in the publication search. Finally, a limitation was the low number of cases in some trial characteristic categories (for example, investigator-initiated trials), which limited the precision of the risk estimates. We could, therefore, not perform multivariable analysis. Conclusions regarding the associations between trial characteristics and discrepancies should be interpreted with caution.

To conclude, protocol-publication discrepancies in clinical drug trials were not unusual in primary endpoints, but common in secondary endpoints, secondary objectives, and subgroup or other additional analyses. Despite the improvement compared to previous studies, the occurrence of discrepancies was still substantial, indicating that selective publication remains a problem in clinical research. Investigator-initiated, not prospectively registered and non-randomized trials were determinants of discrepancies in the primary endpoint. Full transparency of the process of clinical trial protocols to publications can eliminate these opportunities for selective reporting. Practically, this could mean that the original trial protocol and substantial amendments are made publically available at the moment of publication of the results of the trial. That is likely the essential way forward in pursuing ethically sound and scientifically valid clinical research.

SUPPLEMENTARY METHODS AND RESULTS

TABLE S2.4.1 Protocol-publication discrepancies and their categories.

Protocol-publication discrepancies	Discrepancy categories[§]
Endpoints[*]	Primary endpoint discrepant (added, omitted, or changed[‖], and sum total) in publication/no discrepancy in primary, but in secondary endpoint
Objectives[*]	Primary objective discrepant (added, omitted, or changed[‖], and sum total) in publication/no discrepancy in primary, but in secondary objective
Selection criteria	Changed in publication
Sample size[†]	<80% or >120% of sample as estimated in protocol included
Sponsorship acknowledgement	Missing in publication
Subgroup or other additional analyses[‡]	Additional or subgroup analysis discrepant (added, omitted, or changed[‖], and sum total) in publication
Methods used for the data analysis[‡]	Intention-to-treat, per protocol analysis (or other method) changed in publication

* Discrepancies in secondary endpoints or objectives were only scored if there were no discrepancies in the primary endpoints or objectives. This was done because if a trial had already a discrepancy in the primary endpoint, we considered eventual discrepancies in secondary endpoints or objectives as redundant extra information, as selective publication was already at issue due to the discrepancies in primary endpoints/objectives. Discrepancies in secondary endpoints were not subdivided into added/omitted/changed, as these are unlikely to influence the direction of the conclusion of the trial and were therefore considered less important. [†] If a trial contained a dose escalation procedure to find the maximum tolerated dose (MTD), the scoring of discrepancy of sample size is not applicable, as there is no target sample size in such trials. [‡] Only scored among randomized controlled trials, as these aspects are often not a part of non-randomized trials. Randomized exploratory pharmacokinetic/-dynamic, safety/tolerability trials with descriptive objectives were also excluded from this subgroup analysis because of the same reason. [§] For all items, the additional categories not mentioned in the table were no discrepancy or no information provided in protocol/publications. Discrepancy categories were scored as no information provided if the protocol and/or the publication provided insufficient detail to identify the category as being either discrepant or not discrepant. The number of trials with insufficient information is tabulated in table 2. [‖] Added: the publication reported aspects that were not described in the protocol. Omitted: the publication did not report aspects that were described in the protocol. Changed: the publication used another definition than the protocol (for example, a different classification cut-off for a categorical variable), or primary was switched to secondary or vice versa.

TABLE S2.4.2 Table of protocol deviations*.

Description of deviation	When	Why
1. Omission of the double-check by second investigator of 20 additional randomly selected trials	After discussion of the scoring of discrepancies of the first 35 trials by second investigator	Inter-rater agreement was very high, and 100% for the most important discrepancies (primary endpoint)
2. Calculation of overall discrepancy score	After discussion with all other investigators during data collection	Does not add much information
3. Only analyzing determinants of primary endpoint discrepancies, and not the other discrepancies.	After finishing discrepancy scoring	Primary endpoint discrepancies are most likely to influence direction of conclusion. Other discrepancies may be relevant, but for this paper we had to prioritize

* Deviations reported in the table are deviations from the originally published study protocol, included in appendix 2.1.

TABLE S2.4.3 Number of articles found per trial.

Number of article publications found per trial (n = 506)	Number of trials (n = 334)	Cumulative percentage of trials
1	263 (79%)	79%
2	41 (12%)	91%
3	13 (4%)	95%
4	5 (1%)	96%
5	5 (1%)	97%
6	1 (0.3%)	97%
7	2 (0.6%)	98%
9	2 (0.6%)	99%
17	1 (0.3%)	99%
22	1 (0.3%)	100%

TABLE S2.4.4 Number of protocol-publication discrepancies found per trial.

Total number of discrepancies	RCTs (n = 204)			Non-RCTs (n = 130)		
	N	%	Cumulative %	N		Cumulative %
0	34	16.7%	16.7%	44	33.8%	33.8%
1	75	36.8%	53.4%	49	37.7%	71.5%
2	65	31.9%	85.3%	25	19.2%	90.8%
3	23	11.3%	96.6%	6	4.6%	95.4%
4	6	2.9%	99.5%	6	4.6%	100.0%
5	1	0.5%	100.0%	0	0.0%	100.0%

REFERENCES

1. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature.* May 25 2016;533(7604):452-454.

2. Ioannidis JP. Why most published research findings are false. *PLoS Med.* Aug 2005;2(8):e124.

3. Dickersin K, Chalmers I. Recognizing, investigating and dealing with incomplete and biased reporting of clinical research: from Francis Bacon to the WHO. *J R Soc Med.* Dec 2011;104(12):532-538.

4. Bouter LM. Open data are not enough to realize full transparency. *Journal of clinical epidemiology.* Jul 8 2015.

5. Knottnerus JA, Tugwell P. Selection-related bias, an ongoing concern in doing and publishing research. *Journal of clinical epidemiology.* Oct 2014;67(10):1057-1058.

6. Strech D. Normative arguments and new solutions for the unbiased registration and publication of clinical trials. *Journal of clinical epidemiology.* Mar 2012;65(3):276-281.

7. van den Bogert CA, Souverein PC, Brekelmans CT, et al. Occurrence and determinants of selective reporting of clinical drug trials: design of an inception cohort study. *BMJ Open.* 2015;5(7):e007827.

8. Chan AW, Song F, Vickers A, et al. Increasing value and reducing waste: addressing inaccessible research. *Lancet.* Jan 18 2014;383(9913):257-266.

9. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA.* May 26 2004;291(20):2457-2465.

10. Berendt L, Callreus T, Petersen LG, Bach KF, Poulsen HE, Dalhoff K. From protocol to published report: a study of consistency in the reporting of academic drug trials. *Trials.* 2016;17:100.

11. Kasenda B, Schandelmaier S, Sun X, et al. Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications. *BMJ.* 2014;349:g4539.

12. Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA.* Sep 2 2009;302(9):977-984.

13. Raghav KP, Mahajan S, Yao JC, et al. From Protocols to Publications: A Study in Selective Reporting of Outcomes in Randomized Trials in Oncology. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* Nov 1 2015;33(31):3583-3590.

14. Merck Guidelines for Publication of Clinical Trials in the Scientific Literature. https://www.merck.com/research/discovery-and-development/clinical-development/Merck-Guidelines-for-Publication-of-Clinical-Trials-in-the-Scientific-Literature.pdf. Accessed on 16-02-2016.

15. Food and Drug Administration Amendments Act of 2007, Title VIII - Clinical trial databases. Public Law 110–85. September 27, 2007.

16. Amendment of the Medical Research Involving Human Subjects Act with regard to the evaluation of the act and recovery of incomplete implementation of guideline no. 2001/20/EG. Dossier 31452. Enacted as per 1 July 2012 (full text in Dutch only).

17. Trans-NIH BioMedical Informatics Coordinating Committee (BMIC). NIH Data Sharing Repositories. https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html. Last accessed on 18 October 2016.

18. European Federation of Pharmaceutical Industries and Associations (EFPIA) Position - Transparency of Information on Clinical Trials included in the Proposed EU Database (Article 78 of the Proposal for a Regulation on Clinical Trials). http://www.efpia.eu/ uploads/Modules/Documents/efpia_key_messages_transparency_final_10.12.2012-updated_0.pdf. Last accessed on 2 November 2016.

19. ICMJE. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. Update December 2015. http://www.icmje.org/icmje-recommendations.pdf Last accessed on 2 November 2016.

20. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340:c332.

21. van den Bogert CA, Souverein PC, Brekelmans CT, et al. Non-Publication Is Common among Phase 1, Single-Center, Not Prospectively Registered, or Early Terminated Clinical Drug Trials. *PLoS One.* 2016;11(12):e0167709.

22. The COMPare Trials Project. Goldacre B, Drysdale H, Powell-Smith A, et al. www.COMPare-trials.org, 2016. Website last accessed on 18 October 2016.

23. Chan AW, Hrobjartsson A, Jorgensen KJ, Gotzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ.* 2008;337:a2299.

24. Chan AW, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med.* Feb 5 2013;158(3):200-207.

25. Chan AW, Tetzlaff JM, Altman DG, Dickersin K, Moher D. SPIRIT 2013: new guidance for content of clinical trial protocols. *Lancet.* Jan 12 2013;381(9861):91-92.

26. Deangelis CD, Fontanarosa PB. Strengthening the credibility of clinical research. *Lancet.* Jul 24 2010;376(9737):234.

27. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *BMJ.* Oct 24 2016;355:i5239.

28. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA.* Dec 26 2012;308(24):2594-2604.

29. Schulz KF, Moher D, Altman DG. CONSORT 2010 comments. *Lancet.* Oct 9 2010;376(9748):1222-1223.

30. Sune P, Sune JM, Montoro JB. Positive outcomes influence the rate and time to publication, but not the impact factor of publications of clinical trial results. *PLoS One.* 2013;8(1):e54583.

31. Central Committee on Research Involving Human Subjects. Explanation of the General Assessment and Registration (ABR) Form (version 15 December 2015). Available at: http://www.ccmo.nl/attachments/files/b1-abr-toelichting-15-dec-2015-eng.pdf. Last accessed on 30 October 2016.

32. European Medicines Agency. Clinical efficacy and safety guidelines. http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000085.jsp Last accessed on 18 October 2016.

33. Hewitt CE, Mitchell N, Torgerson DJ. Listen to the data when results are not significant. *BMJ.* Jan 5 2008;336(7634):23-25.

34.   Munafo MR, Nosek BA, Bishop DVM, et al. A manifesto for reproducible science. *Nature Human Behaviour.* 2017;1, 0021.

35.   Button KS, Ioannidis JP, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci.* May 2013;14(5):365-376.

36.   van der Valk FM, van Wijk DF, Stroes ES. Serendipity of post-hoc surrogate marker research. *Eur Heart J.* Dec 2012;33(23):2897-2898.

37.   World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* Nov 27 2013;310(20):2191-2194.

38.   Chan AW. Bias, spin, and misreporting: time for full access to trial protocols and results. *PLoS Med.* Nov 25 2008;5(11):e230.

39.   Santori G. Research papers: Journals should drive data reproducibility. *Nature.* Jul 20 2016;535(7612):355.

40.   National Institutes of Health. Principles and Guidelines for Reporting Preclinical Research. https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research. Last accessed on 19 October 2016.

41.   Knottnerus JA, Tugwell P, Tricco AC. Individual patients are the primary source and the target of clinical research. *Journal of clinical epidemiology.* Aug 2016;76:1-3.

42.   Koletsi D, Valla K, Fleming PS, Chaimani A, Pandis N. Assessment of publication bias required improvement in oral health systematic reviews. *Journal of clinical epidemiology.* Aug 2016;76:118-124.

2

Chapter 2.5

One third of clinical drug trials end up in the drug licensing dossier

Cornelis A van den Bogert, Susan W J Janssen, Peter van Meer, Ylfa van Bergen, Patrick C Souverein, Cecile T M Brekelmans, Gerard H Koëter, Hubert G M Leufkens

ABSTRACT

We investigated which clinical drug trials conducted are included in the product dossier of the regulatory authorities among a cohort consisting of all clinical drug trials that were reviewed by the accredited Independent Review Boards in the Netherlands in 2007. Of the analyzed 558 trials, 197 (35.3%) were included in a product dossier. Hundred eighty-three trials were not included because the manufacturer was not licensed in the Netherlands, 153 were not included because not sponsored by the manufacturer, 14 because the indication/dose was not licensed for the product, and 11 because the information was in line with the already licensed claims in the dossier. Thus, all legally required data was submitted to the regulatory product dossier. Findings of trials not sponsored by the manufacturer may, however, be relevant to update the licensed drug label. Incorporating investigator-initiated trials in the regulatory dialogue could optimize the dissemination of clinically relevant findings.

2

INTRODUCTION

The overall purpose of clinical drug trials is to seek answers to relevant questions about the efficacy and safety of the investigated drugs. Based on the findings of trials, drug marketing regulatory authorities (the Medicines Evaluation Board (MEB) in the Netherlands, and the European Medicines Agency (EMA) for the EU) evaluate the risks and benefits for both new and existing drug products on the market, and review whether the submitted evidence can sustain the medical claims made in the drug label. The marketing authorization holders are required to submit all relevant information obtained through clinical trials for inclusion in the marketing authorization product dossier when applying for marketing authorization or post-marketing variations.

A proportion of all clinical drug trials may be specifically initiated with the intention to add information to the clinical data package of the regulatory product dossier. Other drug trials are not initiated with this purpose, but are purely conducted to answer a clinical and/or pharmacological research question and add knowledge to the biomedical literature and clinical treatment guidelines. Results generated by this latter type of drug trial may or may not be required and/or relevant for the regulatory product dossier. Currently, only manufacturers of drug products are authorized to submit data to be incorporated in the claims in the licensed drug label.

Understanding which drug trials are (not) involved in regulatory licensing activities, and why, will inform regulators if all required clinical data is included in the product dossier, and inform prescribers on the extent to which the licensed drug label reflects the actual body of clinical evidence that is available. Therefore, in this study we aimed to identify which clinical drug trials were submitted to the marketing authorization authorities 9-10 years after institutional review board (IRB)-review, among a complete cohort of clinical drug trials within the Netherlands.

METHODS

The design, setting and previous results of this cohort study have been published before[1-4]. Here, a follow-up study of the same dataset was performed, but with a different outcome. In short, we included all clinical drug trials reviewed by the Dutch Institutional Review Boards (IRBs) in 2007 and obtained characteristics of the trials from the electronic clinical trial application form. This form is submitted and saved in the online portal ToetsingOnline, which is required by Dutch

law. Thus, this data source provided a complete selection of all clinical drug trials conducted in the Netherlands. In December 2015, we determined for all trials in the cohort whether they had started and were either completed, discontinued or still running, and whether the results were published as peer-reviewed article in the scientific literature[2]. Figure 2.5.1 shows the selection of the cohort, exclusion of trials that were rejected by the IRB, never started, were still running, and did not involve a drug product eligible for marketing authorization by the MEB or the EMA (for example, transplantation trials, medical device trials, and trials investigating illicit drugs were excluded).
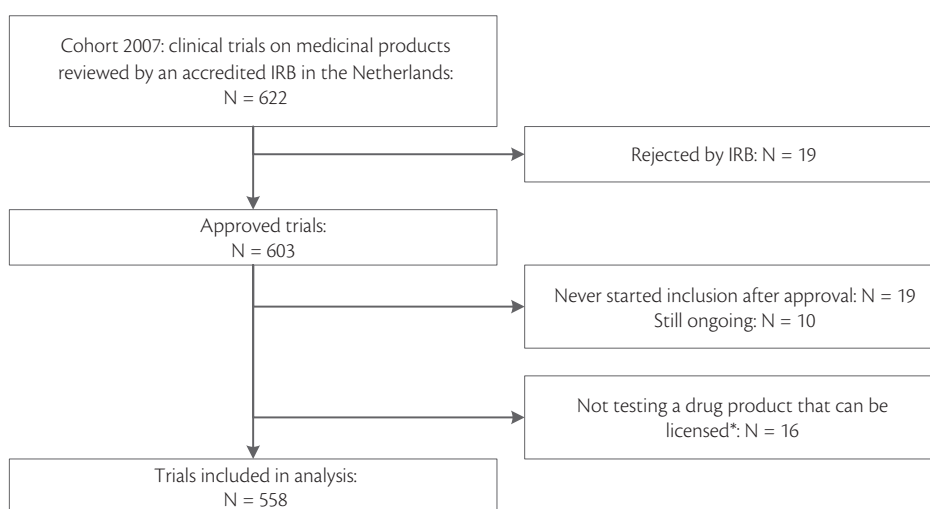

FIGURE 2.5.1 Selection of trials included in the analysis. *Sixteen trials were excluded because they tested a product or therapy that does not qualify for marketing authorization (e.g. medical devices, tissue/cell transplantation, party drugs).

The outcome of interest was the inclusion of the results of the trials in a product dossier of the EMA or MEB. The dossiers of products that were submitted for marketing authorization, either approved or declined by the EMA or the MEB, are since 2007 stored in the Information and Communication Infrastructure (ICI) database.

For each trial, we looked which drug product was stated as the main interest in the trial rationale, objective, and/or hypothesis, and defined this as the "index product". If multiple drugs were stated as the main interest, it could have multiple index products. We classified whether the index product was an innovator or a generic/biosimilar product. The index product was considered as innovator if it was the first product to be licensed with the active substance (the fifth level of the Anatomical Therapeutic Chemical (ATC) classification). Furthermore, we

looked at the global age of the index product. Of drugs that were licensed in the Netherlands, the year that they have become available on any national or international market is available in the drug monograph of the *Informatorium Medicamentorum* (a reference book with monographs of all drugs in use in the Netherlands[5]). Of drugs that were not licensed in the Netherlands and therefore not included in the *Informatorium Medicamentorum*, we used websites of regulatory authorities in the EU and the US, and PubMed, to find out whether they were new or old in the global context. We categorized the global age as not on the market (yet), on the global market for 0-10 years, and on the global market for >10 years.

We searched whether the results of the trials were available in the product dossier. Sources were the licensed drug label (also known as the Summary of Product Characteristics, or SPC), European Public Assessment Report (EPAR), and in ICI. SPCs and EPARs were manually examined to investigate whether the trial was mentioned in the clinical sections. The ICI search was performed with the product names, EudraCT-numbers and the trial protocol numbers as keywords. Trial results were considered as being included in the dossier if a report of the trial was found in the SPC, EPAR, and/or ICI. Possible formats of the results could be full clinical study reports (CSRs), summaries, or copies of literature articles. Reports of single cases of adverse events that occurred in the trials were not considered included if there was no integral description of the results of the trial. One investigator (YB) performed the search for the availability of the results. Then, a second investigator (CAB) double-checked all trials of which the first investigator found no results. The search for inclusion was performed between August and December 2016 (9-10 years after IRB-approval).

For the non-included trials, we investigated why they were not included. Putative reasons for non-inclusion that would be legally justified were assigned in a logical order. The flowchart in figure 2.5.2 shows the order in which all non-included were assigned the reason for non-inclusion. The first possibility was that the index product had no marketing authorization application in the Netherlands. In that case, no product dossier was submitted. The second possibility was that the trial sponsor was not the MAH of the index product in the trial (investigator-initiated trials, or trials sponsored by a company who is not the MAH). Investigator-initiated trials can be carried out with no involvement of the MAH. Only the MAH is allowed in the current system to submit the data to support the therapeutic claims in the product dossier. Thus, investigator-initiated trials can remain outside the scope of the product dossier if the MAH does not include them in a new application. The third possible reason was that the marketing authorization of the index product was not valid for the indication,
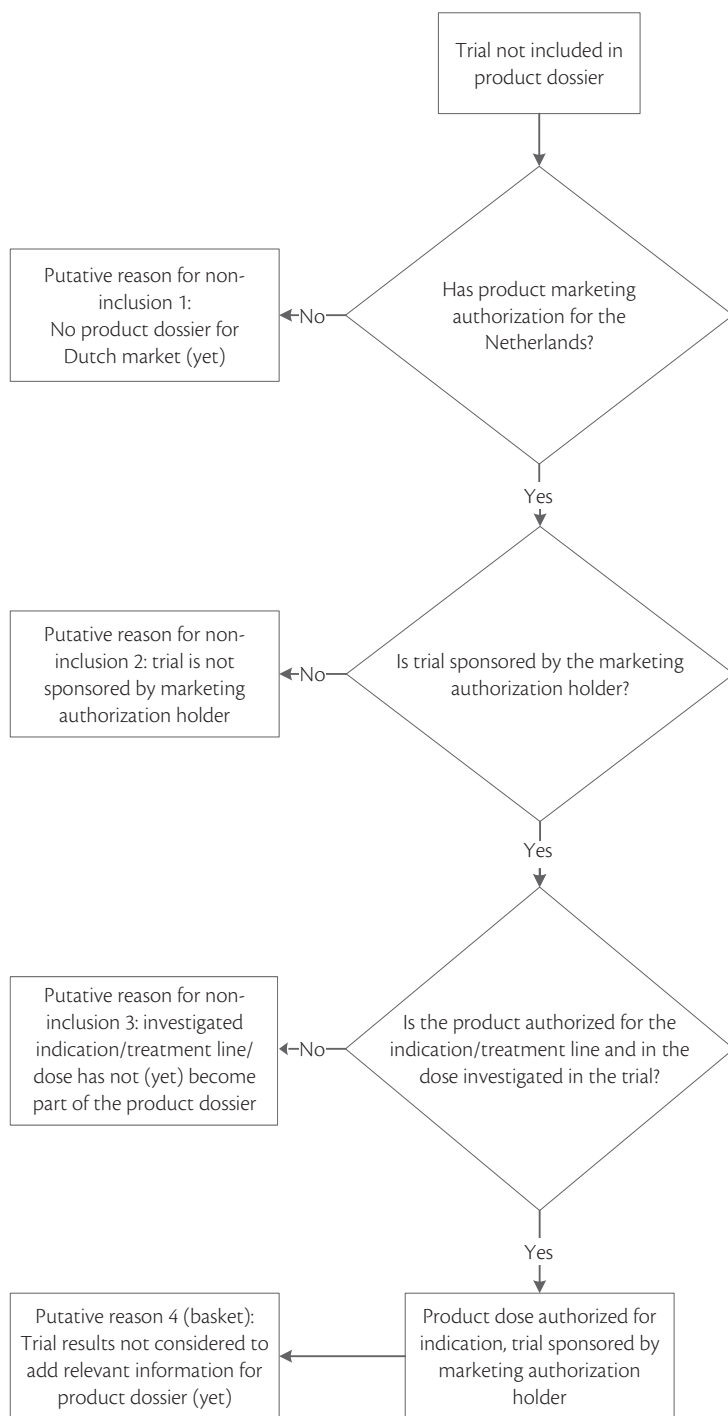
```
                          ┌─────────────────────┐
                          │ Trial not included in│
                          │  product dossier     │
                          └──────────┬──────────┘
                                     │
                                     ▼
                              ╱────────────╲
┌──────────────────────┐     ╱ Has product  ╲
│ Putative reason for   │    ╱  marketing     ╲
│ non-inclusion 1:      │◄─No─  authorization for
│ No product dossier for│    ╲  the Netherlands?╱
│ Dutch market (yet)    │     ╲              ╱
└──────────────────────┘      ╲────────────╱
                                   │ Yes
                                   ▼
                              ╱────────────╲
┌──────────────────────┐     ╱              ╲
│ Putative reason for   │    ╱ Is trial       ╲
│ non-inclusion 2: trial│◄─No─ sponsored by the
│ is not sponsored by   │    ╲ marketing      ╱
│ marketing authorization│    ╲ authorization ╱
│ holder                │      ╲ holder?     ╱
└──────────────────────┘      ╲────────────╱
                                   │ Yes
                                   ▼
                              ╱────────────╲
┌──────────────────────┐     ╱ Is the product╲
│ Putative reason for   │    ╱  authorized for  ╲
│ non-inclusion 3:      │◄─No─ the indication/
│ investigated          │    ╲ treatment line  ╱
│ indication/treatment  │    ╲ and in the dose ╱
│ line/dose has not (yet)│     ╲ investigated  ╱
│ become part of the    │      ╲ in the trial?╱
│ product dossier       │       ╲───────────╱
└──────────────────────┘            │ Yes
                                    ▼
┌──────────────────────┐     ┌──────────────────┐
│ Putative reason 4     │     │ Product dose     │
│ (basket): Trial results│◄────│ authorized for   │
│ not considered to add  │     │ indication, trial│
│ relevant information   │     │ sponsored by     │
│ for product dossier    │     │ marketing        │
│ (yet)                  │     │ authorization    │
└──────────────────────┘     │ holder           │
                             └──────────────────┘
```

FIGURE 2.5.2 Identification of the reason for non-inclusion for each non-included trial.

or the dose, investigated by the trial. Trials could be not included in the dossier because the extension to the indication and/or dosage regimen of the trial was never submitted by the MAH, for example after the trial showed negative results. Finally, if the manufacturer sponsored the trial and the marketing authorization included the combination of the index product and indication, we determined whether the results of the trial were already covered by the current claims and data of the product dossier, identified no issues regarding the product, and were therefore not submitted to the regulatory authorities. We summarized the general objectives of the trials in this fourth category separately.

Trial characteristics were summarized, and the frequencies of included and non-included trials were described for all characteristics. Furthermore, we analyzed the external validity of our cohort by comparing the inclusion percentages of the subgroup of industry-sponsored trials by phase to attrition rates reported in the literature [6,7]. The reasons for non-inclusion explained the differences in inclusion percentages between trial characteristic categories. Kaplan-Meier survival analysis was used to visualize the inclusion of trials over time, starting from IRB-review. The mean and interquartile range (IQR) of the time-to-inclusion were calculated. All calculations were performed in IBM SPSS, version 23 and 24. No ethical approval was required for this study.

RESULTS

Baseline characteristics of the 558 included trials are shown in table 1. Forty-eight percent of the analyzed trials included drugs of which the innovator product has been on the international market for more than 10 years. Of these, we found that a report of the results of 197 trials (35.3%) was included in the product dossier. Results of fifty-eight trials (10.4%) were submitted as part of a marketing authorization for a new product; 60 (10.8%) were submitted to update the drug label (for example, addition of a new indication, or change in treatment line); 56 (10.0%) were submitted as part of a post-marketing surveillance program; 18 (3.2%) as part of pediatric drug development, and 5 (0.8%) as a notification that was not part of a specific regulatory procedure or application.

TABLE 2.5.1 Characteristics of the 558 included trials.

Total clinical trials in cohort	N = 558 (100%)
Characteristics	
Sponsor	
Industry sponsored	348 (62.4%)
Investigator-initiated	210 (37.6%)
Phase	
Phase 1	117 (21.0%)
Phase 2	123 (22.0%)
Phase 3	171 (30.6%)
Phase 4	56 (10.0%)
Other than phase 1-4*	91 (16.3%)
Centers	
Single center	239 (42.8%)
Multicenter(only the Netherlands)	51 (9.1%)
Multiccenter (only the Netherlands and within the EU)	80 (14.3%)
Multicenter (the Netherlands and also outside the EU)	188 (33.7%)
Disease area	
Oncology	112 (20.1%)
Neurological diseases (including analgesia and anaesthesia trials)	72 (12.9%)
Cardiovascular diseases	59 (10.6%)
Endocrine diseases	58 (10.4%)
Infectious diseases (including vaccine trials)	40 (7.2%)
Mental and behavioral disorders	35 (6.3%)
Respiratory diseases	34 (6.1%)
Haematological and immunological diseases	33 (5.9%)
Musculoskeletal diseases	30 (5.4%)
Digestive system diseases	22 (3.9%)
Genitourinary system diseases	24 (4.3%)
Other disease areas	39 (7.0%)
Completion	
Completed as planned	458 (82.1%)
Discontinued	100 (17.9%)
Innovator or generic	
Innovator product	471 (84.4%)
Generic, biosimilar or new formulation	87 (15.6%)

TABLE 2.5.1 Characteristics of the 558 included trials. (continued)

Global age of active substance of index product	
Not on market (yet)	165 (29.6%)
0-10 years	124 (22.2%)
>10 years	269 (48.2%)
Published in the literature as peer-reviewed article	
Published	324 (58.1%)
Not published	234 (41.9%)

Table 2 summarizes the inclusion and reason-specific non-inclusion percentages. Out of 558, 183 trials (32.8%) were not included because the product was not licensed within the Netherlands and hence did not have a product dossier in the database of the regulatory authority. One-hundred fifty-three (27.4%) trials involved a product that was licensed, but the sponsor of these trials was not the marketing authorization holder. These trials were mostly investigator-initiated, with the exception of two industry-sponsored trials that investigated products of which they were not the marketing authorization holder. Fourteen trials (2.5%) sponsored by the marketing authorization holder of the product were not included because the new indication, treatment line or dose had not been submitted by them to the regulatory dossier, for example because the results were negative. Finally, eleven trials (2.0%) sponsored by the marketing authorization holder of the product that was approved for the tested indication and dose were not included. Based on the objectives of these trials, we determined that ten of them were carried out for other purposes than for inclusion in the regulatory product dossier. One trial was testing the efficacy and safety of a product. The results of this trial were published in the scientific literature. We determined that the results and conclusions of this publication were in line with the licensed drug label by the time that the trial was completed. Therefore, updating the product dossier due to this trial was not warranted. The MAH conducted this trial to confirm the efficacy and safety of the product in a subpopulation that had been underrepresented in previous trials.

TABLE 2.5.2 Overview of the inclusion percentage and why trials were not included in the regulatory product dosser.

Total	558 (100%)
Trials included in product dosser	197 (35.5%)
Trials not included in product dossier:	
Index product has no marketing authorization in the Netherlands	183 (32.8%)
Trial is not sponsored by MAH of index product	153 (27.4%)
Index product is sponsored by MAH, no marketing authorization for indication/dose combination	14 (2.5%)
Index product is sponsored by MAH and authorized for dose/indication:	
Biomarker trial	3 (0.5%)
Trial on dose optimization (in combination with other drug therapies)	3 (0.5%)
Trial on difference in patient satisfaction between 2 authorized dosage forms	2 (0.4%)
Trial is conducted as basis for development of new product with similar properties	1 (0.2%)
Trial testing a different non-active ingredient of the drug product	1 (0.2%)
Trial testing efficacy and safety of index product; results in line with dossier	1 (0.2%)

Abbreviation: MAH = marketing authorization holder

Table S2.5.1 (supplementary results) summarizes the percentages of included and non-included trial (per reason), stratified by trial characteristics. Differences in inclusion and non-inclusion percentages per characteristics could be explained by the proportion of investigator-initiated trials in that category. For example, trials on older drugs (>10 years on the international market) were generally less likely to be included in a product dossier compared to trials on newer drugs (0-10 years on the international market) because most of these trials were investigator-initiated. Of the 324 published trials, 158 (48.8%) were not included, vs. 31 (13%) of the 234 unpublished trials. Most of the non-included unpublished trials were not included because the product was not authorized (47.9%), or because they were not sponsored by the MAH (33.3%).

Figure 2.5.3 shows that, based on visual inspection, the inclusion over time was not saturated at the end of follow-up for all trial phases, as new inclusions were observed close to the end of follow-up. The overall median time to inclusion was 4.5 years (IQR 3.0-6.4). Among the subgroup of industry-sponsored phase 1-3 trials, 13 out of the 104 (12.5%) phase 1 trials, 34 out of the 80 (42.5%) phase 2, and 98 out of 138 (71.0%) phase 3 trials were included in the dossier.
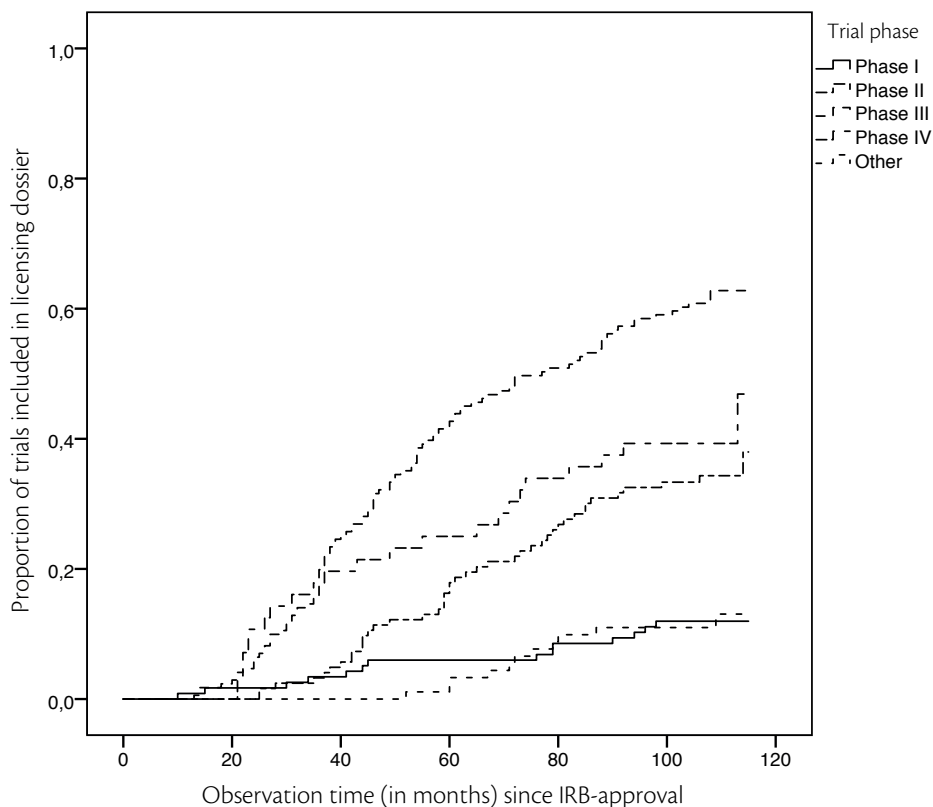
FIGURE 2.5.3 Kaplan-Meier curve of the inclusion of trials in the product dossier stratified by trial phase. T_0 = date of IRB-approval. The median time to inclusion of phase 1 trials was 5.0 years (IQR 2.8-7.6), for the phase 2 trials also 5.0 years (IQR 3.7-6.7), for the phase 3 trials 3.8 years (IQR 2.9-5.6), for the phase 4 trials 3.6 years (IQR 1.9-7.3), and for the trials other than phase 1-4 6.0 years (IQR 5.0-6.7).

DISCUSSION

We found that after 10 years following IRB approval of the trial protocol, 35% of the clinical drug trials approved by Dutch IRBs were included in the regulatory product dossier. Our analysis of the non-included trials demonstrated that all trials that were required to support the licensed label claims of the drug products at the time of application of the license, were included in the dossier. Non-included drug trials involved unlicensed products, investigator-initiated trials, trials testing dosages or indications that were not licensed, or confirmed the benefit/risk profile for subpopulations within the licensed indication.

Most non-included trials were sponsored by the manufacturer, testing products that have no marketing authorization for the drug product, or testing drugs

against indications in dosages that were not included in the licensed drug label. These categories can be considered as attrition in clinical drug development: the product, indication and/or dosage did not prove efficacious and therefore, no licensing follow-up activity occurred (yet). According to two previous studies[6,7], the probability that a phase 1 trial leads to a marketing authorization of a new drug is 10-32%. This probability is 16-38% for phase 2 trials and 50-71% for phase 3 trials. The inclusion percentages of the industry-sponsored phase 1-3 trials were within these ranges of attrition rates in clinical drug development. The disease areas were represented in approximately similar proportions as in studies based on global clinical trial databases. These findings suggest that our national cohort is indeed a good representation of clinical drug development in the international context.

Twenty-seven percent of the trials were not included in the product dossier because they were not sponsored by the manufacturer of the drug product, but (mostly) investigator-initiated. This is an interesting category, as these trials are likely not intended to be used for licensing, but the results of these trials might nevertheless be relevant for the medical claims stated in the licensed drug label. This is not necessarily problematic, as if these trials disseminate useful therapeutic findings through scientific publication, they will supposedly be incorporated in the clinical treatment guideline update. Moreover, if worrying safety signals emerge from such trials, alternative regulatory pathways exist to quickly inform prescribers and patients [8,9] and to update the licensed drug label as well. However, there may be cases where such pathways for investigator-initiated trials are also warranted for efficacy-related findings, especially if subgroups are identified for which a product is not, or less efficacious. Guideline updates can take several years, and meanwhile patients should not be treated with drugs that investigator-initiated trials demonstrated to be inefficacious. In the current system, the manufacturer of the drug product is the only authorized stakeholder to submit data to the regulatory product dossier. It is not in the interest of the manufacturer to quickly apply for a restricting the licensed indications of its product. Therefore, to avoid delay and prejudice, submission of contradictory trial results to the regulatory authorities can better be done directly by the investigators of the trial. An investigator-initiated trial in France, for example, demonstrated that epidermal growth factor receptor inhibitors were ineffective against KRAS-mutated colorectal cancer[10]. The claim in in the licensed drug label of cetuximab at that time did not exclude patients carrying this mutation. The investigator-initiated trial was published in February 2007, but the licensed drug label was not restricted until September 2008 [11]. Prescribers and patients unaware of this publication and faring on the medical claims of the drug label

would have made different treatment decisions and, likely, could have had better outcomes.

Coles & Cloyd described the role of academic institutions in neglected diseases and orphan drugs [12]. They acknowledged the importance of investigator-initiated research, but also identified several challenges including lack of infrastructure, regulatory experience, funding, scientific career-related issues, globalization, collaboration with the industry and conflicts of interest. We concur that these practicalities need to be solved before investigator-initiated data can be fully integrated in the licensing dialogue. Notwithstanding these challenges, in the pursuit of adaptive pathways to drug licensing, investigator-initiated research represents a valuable addition to the proposed adaptations [13-15].

Our study has some limitations. First, the question can be raised how representative our cohort was, as regulatory product dossiers consist of data collected all over the world [16], illustrated by figure 2.5.4. As some trials were only carried out in the Netherlands, some trials will also only be carried out in Germany, United Kingdom, United States, etc., and these were therefore not included in our dataset. Over 60% of patients involved in pivotal trials supporting marketing authorization applications to the EMA came from countries outside the EU [16]. In our cohort, 122 trials (21.9%) were multicenter and conducted in countries outside the EU. This indicates that product dossiers contain also many trials not conducted in the Netherlands. In addition, the trials analyzed in our cohort may be included in product dossiers of regulatory authorities elsewhere. Secondly, the follow-up from IRB-review was 9-10 years, which may be in some cases too short to definitely conclude whether the trial is included in the dossier. The percentage of 35% should therefore be interpreted with caution – in particular, the early-phase trials may be part of a development program that will lead to a license dossier in the future. The second limitation is that we had only access to the database of the Dutch regulatory authority, which contains the product dossiers of drugs that have an EU-wide license or a license only for the Netherlands. Therefore, trials that were not included in the product dossier because the drug was not licensed in the Netherlands could still be included in product dossiers elsewhere.

In conclusion, one-third of all clinical trials conducted in the Netherlands were included in a product dossier of the drug marketing regulatory authorities, 9-10 years after follow-up. All information that should have been in the product dossier according to the legal requirements was included. The two main reasons for non-inclusion were no existing product dossier in the Netherlands and that the trial was investigator-initiated with no intention to be included the drug licensing dossier. Investigator-initiated trials could be used more often in the future to update the regulatory drug label efficiently and timely.
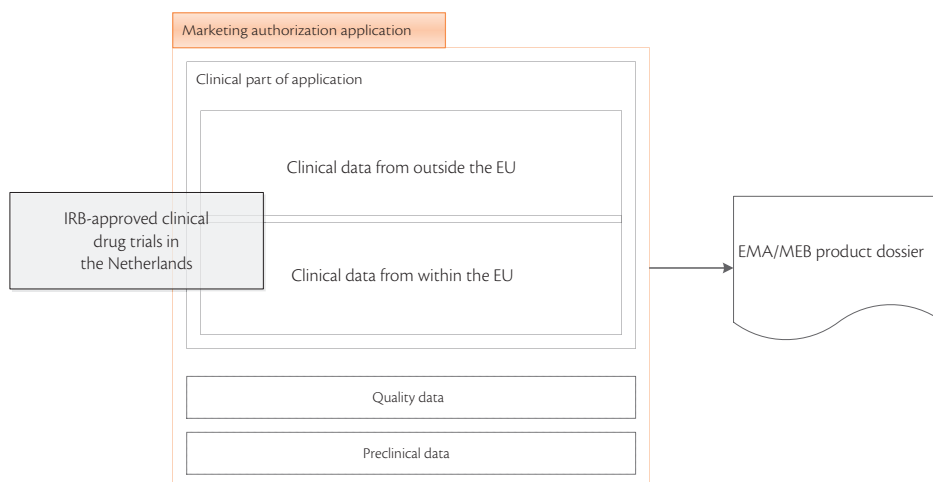
2

FIGURE 2.5.4 Schematic overview of the role of IRB-approved clinical drug trials in the Netherlands in the composition of regulatory product dossiers. The trial activity in the Netherlands consists partially of international multicenter trials (mostly within, and also outside the EU), and partially of trials only conducted within the country. The clinical part of marketing authorization application also contains trials that were not conducted in the Netherlands.

SUPPLEMENTARY RESULTS

TABLE S2.5.1 Percentages of included and non-included trials (per reason), summarized per trial characteristic

	N (%) included	N (%) not included (product not authorized)	N (%) not included (trial not sponsored)	N (%) not included (indication/ dose not authorized)	N (%) not included (no new information)
Sponsor					
Industry sponsored (n = 348)	163 (46.8%)	158 (45.4%)	2 (0.6%)	14 (4.0%)	11 (3.2%)
Investigator-initiated (n = 210)	34 (16.2%)	25 (11.9%)	151 (71.9%)	0 (0.0%)	0 (0.0%)
Phase					
Phase 1 (n = 117)	14 (12.0%)	90 (76.9%)	9 (7.7%)	2 (1.7%)	2 (1.7%)
Phase 2 (n = 123)	43 (35.0%)	41 (33.3%)	34 (27.6%)	5 (4.1%)	0 (0.0%)
Phase 3 (n = 171)	106 (62.0%)	33 (19.3%)	23 (13.5%)	4 (2.3%)	5 (2.9%)
Phase 4 (n = 56)	23 (41.1%)	5 (8.9%)	24 (42.9%)	2 (3.6%)	2 (3.6%)
Other than phase 1-4* (n = 91)	11 (12.1%)	14 (15.4%)	63 (69.2%)	1 (1.1%)	2 (2.2%)
Centers					
Single center (n = 239)	30 (12.6%)	98 (41.0%)	107 (44.8%)	1 (0.4%)	3 (1.3%)
Multicenter (only the Netherlands)	10 (19.6%)	7 (13.7%)	31 (60.8%)	1 (2.0%)	2 (3.9%)
Multicenter (only the Netherlands and within the EU)	35 (43.8%)	29 (36.3%)	9 (11.3%)	4 (5.0%)	3 (3.8%)
Multicenter (the Netherlands and also outside the EU)	122 (64.9%)	49 (26.1%)	6 (3.2%)	8 (4.3%)	3 (1.6%)
Disease area					
Oncology (n = 112)	44 (39.3%)	28 (25.0%)	36 (32.1%)	4 (3.6%)	0 (0.0%)
Neurological diseases (n = 72)	24 (33.3%)	31 (43.1%)	17 (23.6%)	0 (0.0%)	0 (0.0%)
Cardiovascular diseases (n = 59)	21 (35.6%)	18 (30.5%)	19 (32.2%)	1 (1.7%)	0 (0.0%)
Endocrine diseases (n = 58)	13 (22.4%)	31 (53.4%)	10 (17.2%)	2 (3.4%)	2 (3.4%)
Infectious diseases (n = 40)	21 (52.5%)	15 (37.5%)	3 (7.5%)	1 (2.5%)	0 (0.0%)
Mental and behavioral disorders (n = 35)	4 (11.4%)	13 (37.1%)	16 (45.7%)	2 (5.7%)	0 (0.0%)
Respiratory diseases (n = 34)	13 (38.2%)	17 (50.0%)	1 (2.9%)	1 (2.9%)	2 (5.9%)
Hematological and immunological diseases (n = 33)	15 (45.5%)	8 (24.2%)	9 (27.3%)	1 (3.0%)	0 (0.0%)
Musculoskeletal diseases (n = 30)	9 (30.0%)	6 (20.0%)	10 (33.3%)	1 (3.3%)	4 (13.3%)
Digestive system diseases (n = 22)	12 (54.5%)	5 (22.7%)	5 (22.7%)	0 (0.0%)	0 (0.0%)
Genitourinary system diseases (n = 24)	7 (29.2%)	5 (20.8%)	9 (37.5%)	1 (4.2%)	2 (8.3%)
Other disease areas (n = 39)	14 (35.9%)	6 (15.4%)	18 (46.2%)	0 (0.0%)	1 (2.6%)

TABLE S2.5.1 Percentages of included and non-included trials (per reason), summarized per trial characteristic (continued)

	N (%) included	N (%) not included (product not authorized)	N (%) not included (trial not sponsored)	N (%) not included (indication/ dose not authorized)	N (%) not included (no new information)
Completion					
Completed as planned (n = 458)	176 (38.4%)	144 (31.4%)	119 (26.0%)	10 (2.2%)	9 (2.0%)
Discontinued (n = 100)	21 (21.0%)	39 (39.0%)	34 (34.0%)	4 (4.0%)	2 (2.0%)
Innovator or generic product					
Innovator (n = 471)	176 (37.4%)	169 (35.9%)	106 (22.5%)	11 (2.3%)	9 (1.9%)
Generic (n = 87)	21 (24.1%)	14 (16.1%)	47 (54.0%)	3 (3.4%)	2 (2.3%)
Time since innovator version of index drug has been on the market					
Not on market (yet) (n = 165)	14 (8.5%)	151 (91.5%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
0-10 years (n = 124)	103 (83.1%)	7 (5.6%)	11 (8.9%)	2 (1.6%)	1 (0.8%)
>10 years (n = 269)	80 (29.7%)	25 (9.3%)	142 (52.8%)	12 (4.5%)	10 (3.7%)
Published in the literature					
Published	166 (51.2%)	71 (21.9%)	75 (23.1%)	7 (2.2%)	5 (1.5%)
Not published	31 (13.2%)	112 (47.9%)	78 (33.3%)	7 (3.0%)	6 (2.6%)

2

REFERENCES

1. Van den Bogert CA, Souverein PC, Brekelmans CT, et al. Occurrence and determinants of selective reporting of clinical drug trials: design of an inception cohort study. *BMJ Open.* 2015;5(7):e007827.

2. Van den Bogert CA, Souverein PC, Brekelmans CT, et al. Non-Publication Is Common among Phase 1, Single-Center, Not Prospectively Registered, or Early Terminated Clinical Drug Trials. *PLoS One.* 2016;11(12):e0167709.

3. van den Bogert CA, Souverein PC, Brekelmans CTM, et al. Primary endpoint discrepancies were found in one in ten clinical drug trials. Results of an inception cohort study. *J Clin Epidemiol.* May 20 2017.

4. van den Bogert CA, Souverein PC, Brekelmans CTM, et al. Recruitment failure and futility were the most common reasons for discontinuation of clinical drug trials. Results of a nationwide inception cohort study in the Netherlands. *J Clin Epidemiol.* May 06 2017.

5. Informatorium Medicamentorum (Drug Monographs). The Hague: Royal Dutch Pharmaceutical Association (KNMP); 2017.

6. DiMasi JA, Feldman L, Seckler A, Wilson A. Trends in risks associated with new drug development: success rates for investigational drugs. *Clin Pharmacol Ther.* Mar 2010;87(3):272-277.

7. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. *Nat Biotechnol.* Jan 2014;32(1):40-51.

8. European Medicines Agency. Guideline on good pharmacovigilance practices (GVP) Module XV – Safety communication. 26 July 2012. EMA/118465/2012. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/07/WC500130396.pdf. Accessed 23 March 2017.

9. Organisation for Economic Co-operation and Development., Inter-Organization Programme for the Sound Management of Chemicals. *Guidance on safety performance indicators : guidance for industry, public authorities and communities for developing SPI programmes related to chemical accident prevention, preparedness and response.* Paris: OECD; 2003.

10. Di Fiore F, Blanchard F, Charbonnier F, et al. Clinical relevance of KRAS mutation detection in metastatic colorectal cancer treated by Cetuximab plus chemotherapy. *Br J Cancer.* Apr 23 2007;96(8):1166-1169.

11. Rathore AS, Sofer GK. *Process validation in manufacturing of biopharmaceuticals.* 3rd ed. Boca Raton: Taylor & Francis/CRC Press; 2012.

12. Coles LD, Cloyd JC. The role of academic institutions in the development of drugs for rare and neglected diseases. *Clin Pharmacol Ther.* Aug 2012;92(2):193-202.

13. Eichler HG, Baird LG, Barker R, et al. From adaptive licensing to adaptive pathways: delivering a flexible life-span approach to bring new drugs to patients. *Clin Pharmacol Ther.* Mar 2015;97(3):234-246.

14. Eichler HG, Oye K, Baird LG, et al. Adaptive licensing: taking the next step in the evolution of drug approval. *Clin Pharmacol Ther.* Mar 2012;91(3):426-437.

15. Pignatti F, Ashby D, Brass EP, et al. Structured Frameworks to Increase the Transparency of the Assessment of Benefits and Risks of Medicines: Current Status and Possible Future Directions. *Clin Pharmacol Ther.* Nov 2015;98(5):522-533.

16. Healy MJ, Saracci R. Report of a symposium sponsored by the UICC on principles and practice in clinical trials held at the "Statistical Research Unit of the Institut National de la Sante et de la Recherche Medicale", Villejuif, September 15-17, 1969. *Int J Cancer.* Nov 15 1971;8(3):541-545.

Appendix 2.1

Occurrence and determinants of selective reporting of clinical drug trials: design and characteristics of an inception cohort study

Cornelis A. van den Bogert; Patrick C. Souverein, PhD; Cecile T.M. Brekelmans; Susan W.J. Janssen; Manon van Hunnik; Gerard H. Koëter; Hubertus G.M. Leufkens; Lex M. Bouter

ABSTRACT

Introduction

Responsible conduct of research implies that results of clinical trials should be completely and adequately reported. This article describes the design of a cohort study that aims to investigate the occurrence and the determinants of selective reporting in an inception cohort of all clinical drug trials that were reviewed by the Dutch Institutional Review Boards (IRBs) in 2007. It also describes the characteristics of the study cohort.

Methods and analysis

In 2007, Dutch MRECs reviewed 622 clinical drug trials. For each trial, we assessed the stages of progress. We discriminated five intermediate stages and five definite stages. Intermediate stages of progress are: approved by an IRB; started inclusion; completed as planned; terminated early; published as article. The definite stages of progress are: rejected by an IRB; never started inclusion; not published as article; completely reported; selectively reported.

We will use univariate and multivariate Cox regression models to identify trial characteristics associated with non- publication.

We will identify seven trial-specific discrepancy items including the objectives, inclusion and exclusion criteria, endpoints, sample size, additional analyses, type of population analysis, and sponsor acknowledgement. The percentage of trials with discrepancies between the protocol and the publication will be scored. We will investigate the association between trial characteristics and the occurrence of discrepancies.

INTRODUCTION

Responsible conduct of clinical research implies that results of clinical trials should be completely and adequately reported[1,2]. However, a significant part of clinical trial results is never reported: on average, only 50 percent of clinical trials that are started are published in the scientific literature[3-20]. As reporting often depends on the magnitude or direction of the trial conclusions, incomplete reporting may result in publication bias[8,9,19,21-24]. For example, if negative findings are more often not published than positive findings, overall evidence synthesis will be biased, which can harm patients[25-27].

Publishing negative results is sometimes judged irrelevant or uninteresting by the investigator, the journal editor or the sponsor of the trial[28]. Negative trials, however, add valuable information to the body of evidence on the effects of the interventions studied. Moreover, publishing negative findings can prevent the start of unnecessary new clinical trials. This may make the use of resources for investigators and sponsors more efficient[29,30].

Selective reporting of trial results comes in two forms. Firstly, selective reporting can mean that the trial at issue is never published in the scientific literature (non-publication). This can be judged by searching for publications on trials included in an inception cohort, e.g. using information from a trial register[6,12,16,31]. Secondly, selective reporting may indicate that a trial is published in the scientific literature with changes, additions, or omissions of study aspects or findings (selective publication)[32-34]. This second meaning is more subtle and can only be judged by comparing published reports to the full original study protocol.

Non-publication rates of 10 to 88 percent have been reported in the literature[3,5,7-12,14-19]. Selective publication was identified by studying discrepancies between the protocol and publication in reporting endpoints, sample size, statistical methods and subgroup analysis[33,35-37].

That non-publication and selective publication can lead to patient harm was also shown for clinical trials with drugs intended for marketing authorization[15,38,39]. Some new drugs had to be withdrawn from the market after additional data was revealed, showing harmful effects. For example, clinical data on the new anti-inflammatory drug rofecoxib were neither published in the literature, nor revealed to the regulators[39]. Other examples of non-publication and selective publication potentially resulting in patient harm include the antihypertensive drug reboxetine[38], and the antiarrhythmic drug lorcainide[22]. The negative media attention about these and other drug trials has caused a decrease of the public's trust in the pharmaceutical industry and medical research[40,41]. Since then, various codes and guidelines aiming at reducing selective reporting[42-44] were devel-

oped. However, recent research showed that these guidelines have only reduced selective reporting marginally[45,46].

Most studies that investigate selective reporting use data from a public registry, like clinicaltrials.gov. However, not all clinical trials are registered in public registries, and details of the original trial protocol are often unclear or lacking because these registers often do not include full study protocols. In addition, information published in public registries may be subject to selective reporting as well. The availability of the full and original trial protocol submitted to an Institutional Review Board (IRB) enables to track the stages of progress of a study from the start. Therefore, to our opinion, starting with a series of consecutive full trial protocols submitted to an IRB in a defined time window and in a defined area is the best approach to examine non-publication and selective publication. To date, few studies have been done using this approach[47].

We report the design of a study that aims to evaluate reporting practices in an inception cohort of clinical drug trials in the Netherlands. The primary objective of the study is to investigate non-publication and selective publication in an inception cohort of clinical drug trials. With regard to non-publication, we will identify factors associated with non-publication. With regard to selective publication, we will evaluate factors associated with discrepancies between the protocol and the publications of the trials. The secondary objective of this study is to investigate whether selective publication is associated with the direction of trial conclusions. Furthermore, we describe the characteristics of the study cohort.

METHODS AND ANALYSIS

Characteristics and data sources

We identified all clinical drug trials reviewed by the Dutch accredited IRBs [48] between 1 January 2007 and 31 December 2007 (n = 622). These trials define the inception cohort. According to previous studies, a seven year time window is sufficient for most trials to recruit participants, collect data, prepare a manuscript and publish the manuscript[5,8,16].

In addition, we identified the characteristics of these trials (supplementary methods and results, table SA2.1.1). The used source was the General Assessment and Registration (GAR) form. This is a standard obligatory form that investigators submit to the IRB. For 194 trials, multiple therapeutic areas were indicated. Two investigators (CAB and CTMB) independently examined whether these trials could be reclassified to a single therapeutic area and reclassified the combination trials as one therapeutic area. Differences were solved by consensus

TABLE A2.1.1 Planned analyses, endpoints and determinants.

	Determinants	Endpoint
Analysis of non-publication	Trial characteristics + early termination + prospective registration on clinicaltrials.gov	Publication as peer-reviewed article
Analysis of selective publication*	Trial characteristics + early termination + prospective registration on clinicaltrials.gov	Discrepancies between protocol and publication
	Discrepancies between protocol and publication[†]	Direction of publication conclusions*

*Only among published trials; [†]Only among randomized trials

after involving a third investigator (GHK). To reduce the large number of different therapeutic areas, we reclassified the variable to the International Classification of Diseases, version 10[49]. This reclassification retained 11 therapeutic areas and 1 'other' category.

From the trials included, we will extract data on the stages of progress, non-publication and selective publication. In addition to the public data sources and original trial protocols, we plan to send out a questionnaire to the investigators (see appendix 2.2). An overview of the variables we plan to extract is presented in the data extraction form (supplementary methods and results, table SA2.1.2).

Stages of progress

For the 622 trials in the inception cohort, we will determine the various stages of progress (figure A2.1.1). For each clinical drug trial, we will discriminate ten stages of progress. Of these, five are intermediate (meaning that further action is observed or possible), and five are definite (meaning that no further action is observed or possible). We named the stages of progress according to the flow of the cohort, shown in figure A2.1.1. The intermediate stages of progress are: B1 approved by IRB; C1 started inclusion; D1 completed as planned; D2 terminated early; E1 published as article. The definite stages of progress are: B2 rejected by IRB; C2 Never started inclusion; E2 Not published as article; F1 completely reported; F2 selectively reported. We primarily aim to investigate the publication-related stages of progress E1, E2, F1, and F2. However, to understand why these stages of progress are not reached, we also determine the other stages of progress. The stage of progress F2 (selectively reported) is definite for the end of our data collection; later publications can still fill remaining gaps, moving trials to F1 (completely reported).

```
┌─────────────────────────────────────────────────┐
│ A. Inception cohort: clinical drug trials        │
│    reviewed by a Dutch IRB in 2007               │
│    N = 622                                        │
└─────────────────────────────────────────────────┘
```

B1. Approved by Dutch IRB N = 603	B2. Rejected by Dutch IRB **N = 19**
C1. Started inclusion N =	C2. Never started inclusion N =
D1. Completed as planned N =	D2. prematurely ended N =
E1. Published as article N =	E2. Not published as article N =
F1. Completely reported N =	F2. Selectively reported N =

Reasons for non-publication

Discrepancies: Types Determinants

FIGURE A2.1.1 Overview of stages of progress of the 2007 inception cohort of clinical drug trials in the Netherlands. The numbers in the boxes indicate the numbers of trials that succeeded to the specific stages of progress. From B1, C1, D1, E1 to F1 is the 'perfect' flow of a trial in the cohort, meaning that all aspects took place according to the application. The sum of the boxes B2, C2, E2 F1 and F2, which are the five final stages of progress, will be 622.

Non-publication

We search for publications on the trial results in the scientific literature using a standardized algorithm (figure A2.1.2). A publication is defined as a peer-reviewed article containing at least methods and results. All reports not fulfilling this publication (e.g. results reported in registries, conference abstracts containing results, trial summaries on sponsor websites containing methods and results) will be also collected. Peer-reviewed publication is in our opinion the golden standard for reporting clinical research, but trial results can be reported by other means (e.g. registries, sponsor websites, conference abstracts). Using only peer-reviewed articles as endpoint for non-publication is in line with the majority of other research[47]. If we identify more than one publication of trial results, we classify the publication as either primary (i.e., containing the overall results and conclusions) or secondary (i.e., interim, post hoc, subgroup or other analysis). In general, we assume that this will be clearly stated in the publications[42]. Other information collected includes the full-text of the article, the journal, and the first date of publication (e.g., advance online publication). We have completed this part of the publication search in March 2015.
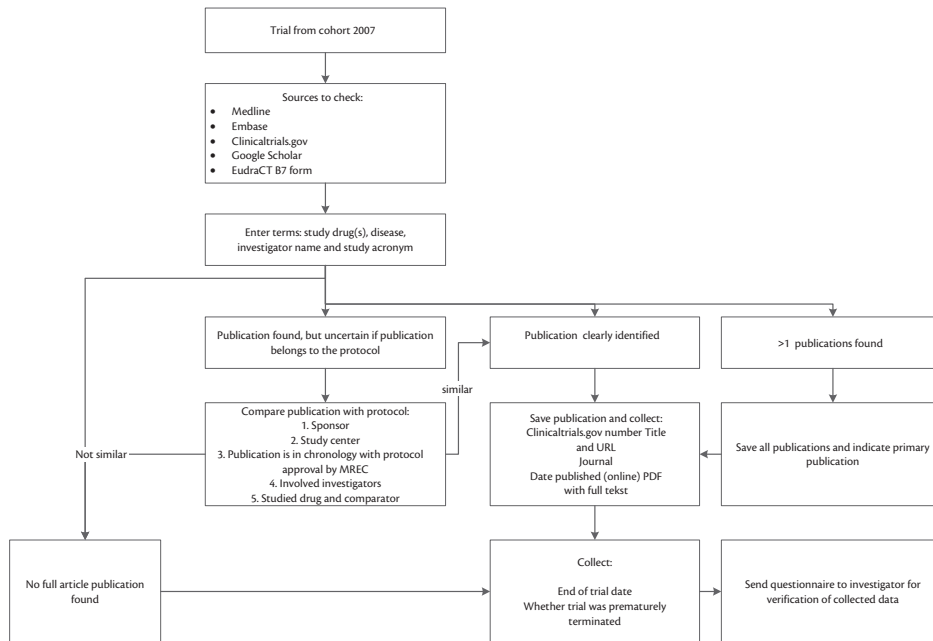
FIGURE A2.1.2 Publication search algorithm. EudraCT = European Union Drug Regulating Authorities Clinical Trials database: obligatory registration database for clinical drug trials carried out in the European Union.

We will also collect the end of trial date and information about (early) termination of the trial. We define the end of trial date as the date of the last visit of the last patient undergoing the trial[50]. A trial is terminated early if either the inclusion or the follow-up is terminated earlier than foreseen in the research protocol. Because early termination is an intermediate stage of progress of a trial, we include early termination as a potential determinant for the endpoints studied. In addition, prospective registration on clinicaltrials.gov will be examined as a potential determinant (table A2.1.1). We define prospective registration as registration of the trial before the first patient is recruited[1]. The data field 'first received' on clinicaltrials.gov will be used as the date of registration.

To validate the used publication search algorithm, two investigators independently searched for publications using the algorithm, using a random selection of 30 trials of the cohort. The two searches identified no differences. We checked the external validity of the algorithm by comparing the results to a search algorithm used for another study[33], kindly provided by the investigators. This comparison showed no differences, which suggested that the construct validity of our algorithm was adequate.

In addition, we will send questionnaires to the main investigators of the research divisions or hospital departments that conducted the trials. We will specifically

TABLE A2.1.2 Reasons for not publishing results, to be obtained from the questionnaire (for unpublished completed trials in cohort).

·   Manuscript is in preparation / under review
·   Results were not interesting enough to publish
·   Journal rejected the manuscript
·   Sponsor decided not to publish without providing a reason
·   Other

ask the investigators to confirm or rectify our information about which stage of progress the trial reached according to our findings. For the non-published trials we ask for the reasons that the trial was not published (table A2.1.2), and whether the results of the trial were reported in alternative ways, such as on clinicaltrials.gov. When the investigator does not respond to the mailed question-naire, we will try to engage the investigator by telephone contact. In case we are unable to contact the investigator, we will contact the sponsor of the trial.

The various stages of progress of the trial in the flowchart will be updated according to the results of the questionnaire. In case neither the investigator nor the sponsor could be reached, the stages of progress remain unchanged. We as-sume that if a trial was incorrectly placed in the stage of progress boxes C2, D2, or E2, the investigator or sponsor would have responded. If we are unable to find any information on whether a trial started inclusion, ended, or was published, we will exclude the trial for subsequent analysis. After showing construct validity, the publication search was performed by two authors (CAB and MH), double-checked by the questionnaire to the investigators. To assess the likelihood of bias, we will investigate whether the characteristics of included cases differ from excluded cases.

Selective publication

Among the trial protocols that resulted in a publication, we will further in-vestigate selective publication. We include only peer-reviewed articles for the discrepancy analysis because other reports contain too little detail to investigate discrepancies with the trial protocol. Selective publication can be measured by identifying discrepancies between protocol and publication. Discrepancies between protocol and publication are indications of selective publication, which may lead to reporting bias. The degree of the risk of reporting bias depends on the association of discrepancies with the direction of trial conclusions. Therefore, among the trials with a randomized design, we will also assess the direction of publication conclusions and investigate whether the direction of publication conclusions is associated with discrepancies between protocol and publication.

2

We define discrepancies between protocol and primary publication as additions, omissions, or changes in pre-specified discrepancy-items. To identify discrepancies systematically, we developed an extraction form containing relevant items. We used items from common protocol and publication guidelines like SIRIT and CONSORT to compel a list with trial items that should be reported. From that list, we selected seven items in which we expected selective reporting (supplementary methods and results, table SA2.1.2)[33,36,37]. The seven discrepancy items include: (1) objectives, (2) inclusion and exclusion criteria, (3) endpoints, (4) sample size, (5) additional analyses, (6) type of population analysis, and (7) sponsor acknowledgement. We will extract these items both from the protocols and the publications. Subsequently, we will compare the extracted data of the protocol to the publications. With regard to discrepancies in the objectives and endpoints, we will distinguish between discrepancies in the primary and in the secondary objectives and endpoints. With regard to discrepancies in the inclusion and exclusion criteria, we will only consider an objective change as discrepancy because inclusion and exclusion criteria are often not fully reported in publications due to the limited availability of space. We will operationalize discrepancies in the planned vs. included sample size as the ratio of sample size achieved divided by sample size planned. With regard to discrepancies in the type of population analysis, we will assess whether an intention to treat or per protocol analysis was planned and used accordingly. We will also indicate when there was a lack of information in the protocol and/or in the publication to assess a discrepancy.

In case we identify multiple publications of one trial protocol, we will include the primary publication in the discrepancy assessment. In addition, if a secondary publication contains any analyses that were not described in the study protocol and this was not stated in the publication, we classify that as an additional discrepancy.

The discrepancy assessment was developed by one author (CAB), and will be tested for construct validity by a second author (PCS), by performing an independent discrepancy assessment of a random selection of 10% of the published trials. Remaining differences will be solved by discussing them with two other authors (CTMB and SWJ). The remaining trials will then be assessed by one author (CAB), with a randomly selected double-check of 20 of the published trials by a second author (PCS). Uncertainties will be solved by a discussion involving two other authors (CTMB and SWJJ).

Among the trials with a randomized design, we will classify the direction of publication conclusions as either positive or negative. This classification is included to investigate whether discrepancies are associated with the direction of

the conclusions (and the interpretation) that the authors draw in the discussion sections of the publications. If trials with a positive conclusion have more discrepancies than trials with a negative conclusion, this may mean that discrepancies are used to spin trial conclusions towards a positive direction. . Two independent investigators (CAB and PCS) will independently classify the trials, and solve differences by consensus.

Data analysis

According to the objectives of the study, we will analyze three endpoints (table A2.1.1): non-publication, discrepancies between the protocol and the publication as a proxy for selective publication, and the direction of publication conclusions.

Non-publication

In a survival analysis of the non-publication rate, only trials that started inclusion will be analyzed (box C1 of figure A2.1.1). The endpoint used is non-publication as peer-reviewed article, according to the definition provided above. The trial end date marks the start of follow-up (i.e. the date the trial transits to the stage of progress D1 or D2, figure A2.1.1). We chose this date instead of the date of IRB approval, because the trials in the cohort might differ in time span. This time span may depend, for example, on the phase of the trial and the number of participants to be recruited. In case of multiple publications of one trial protocol, we use the publication date of the primary publication.

We assume that all trials that started including patients are eligible for publication. Thus, the population of the non-publication survival analysis includes all trials that started inclusion (box C1, figure A2.1.1). Trials that never started inclusion are excluded from this analysis.

To identify characteristics that are associated with (non-)publication, we perform Cox regression analysis to estimate the strength of the association between characteristics and publication status, expressed as hazard ratios and 95% confidence intervals. Because trials of oncolytic drugs are different with respect to the disease severity compared to most trials in other therapeutic areas (which may affect publication), a stratified analysis will be conducted as well. In addition, we will tabulate reasons for non-publication. Finally, we will describe the means of publication by other means than by the definition of publication. By doing so, we will identify the subset of trials with no results reported at all (not as peer-reviewed article and not by other means).

Selective publication

For each of the seven discrepancy-items, we calculate the proportion of trials with the discrepancy. We investigate the association between characteristics and discrepancies for each item (chi-square test) and for the total discrepancy summary score (paired t-test). We will use multivariate logistic (individual discrepancies) and linear (total discrepancy score) regression models to estimate the strength of the association of characteristics and publication status, expressed as odds ratios and 95% confidence intervals. Among the trials with a randomized design, we investigate whether the discrepancies are associated with the direction of the publication conclusions using identical bivariate and multivariate analyses. Data analysis will be performed by two authors (CAB and PCS), and double-checked by all other authors.

By measuring non-publication and selective publication, the study will identify the extent of research underreporting waste in a cohort of clinical trials in the Netherlands[51,52]. To increase the value derived from clinical trials, transparency from protocol to the public is needed[53]. Our study will provide this on a national level and may elucidate areas for improvement. Ultimately, this study may contribute to evidence-based medicine by improving the unbiased reporting rates of clinical drug trials. This may increase the overall trust in research on drugs and the willingness of participants to enroll in clinical drug trials.

SUPPLEMENTARY METHODS AND RESULTS

TABLE SA2.1.1 Cohort 2007 characteristics retrieved from the General Review and Registration-forms.

	N	%
Total clinical trials in cohort	622	100.0%
Sponsor		
Pharmaceutical industry	372	59.8%
Investigator (industry (co-)funded)	74	11.9%
Investigator (no industry funding involved)	176	28.3%
Applicant		
CRO	220	35.4%
Investigator	402	64.6%
Centers involved		
Single center	274	44.1%
Multi center, only in Netherlands	61	9.8%
Multi center, Netherlands and EU	87	14.0%
Multi center, Netherlands and rest of the world	200	32.2%
Phase of study		
Phase 1	125	20.1%
Phase 2	137	22.0%
Phase 3	185	29.7%
Phase 4	66	10.6%
Other/not applicable	109	17.5%
Therapeutic/non therapeutic		
Therapeutic	386	62.1%
Non-therapeutic	236	37.9%
Intervention/observational		
Intervention	556	89.4%
Observational, invasive	51	8.2%
Observational, non-invasive	15	2.4%
Participant category		
≥18 years old and mentally capacitated	571	91.8%
<18 years old and/or mentally incapacitated	51	8.2%
Registration status of product		
Unregistered product	297	47.7%
Registered, studied outside indication	159	25.6%
Registered, studied within indication	128	20.6%
No registration status indicated	38	6.1%

TABLE SA2.1.1 Cohort 2007 characteristics retrieved from the General Review and Registration-forms. (continued)

	N	%
Product category		
Regular medicinal product	590	94.9%
Complex product involved: vaccine, radiopharmaceutical, somatic cell therapy, antisense oligonucleotide	32	5.1%
Therapeutic area		
Neoplasms	117	18.8%
Neurological diseases (including analgesia and anesthesia trials)	74	11.9%
Endocrine diseases	70	11.3%
Cardiovascular diseases	68	10.9%
Mental and behavioral disorders	45	7.2%
Infectious diseases (including vaccine trials)	44	7.1%
Hematological and immunological diseases	38	6.1%
Respiratory diseases	36	5.8%
Musculoskeletal diseases	34	5.5%
Digestive system diseases	26	4.2%
Genitourinary system diseases	25	4.0%
Other	45	7.2%

2

TABLE SA2.1.2 Data extraction form. GAR = General Assessment and Registration

Extract:	Source	Use	If categorical, options
Approved or rejected by Dutch medical research ethics committee	GAR-form	Stage of progress B1/B2	Approved / rejected
Started inclusion/never started inclusion	Questionnaire	Stage of progress C1/C2	Started / never started
Completed as planned/ preliminary terminated	EudraCT B7-form and questionnaire	Stage of progress D1/D2; determinant	Completed as planned / preliminary terminated
End of trial date	EudraCT B7-form and questionnaire	Time to publication calculation	
Publication date	PubMed and questionnaire	Time to publication calculation	Published (yes/no); if yes, date of online publication
Completely reported / selectively reported	Protocol and publication	Stage of progress F1/F2; endpoint	Completely reported / selectively reported
If not published: reason for non-publication	Questionnaire	Reasons for non-publication	
Randomized/non-randomized	Protocol	Characteristic	Randomized / non-randomized
Trial framework	Protocol	Characteristic	Single-arm / parallel group / crossover/ adaptive; superiority / non-inferiority / exploratory / no information*
Primary, secondary, and other/ exploratory objectives	Protocol; publication	Discrepancy-item 1	No discrepancies / primary objectives added / primary objectives omitted / primary objectives changed / other additions, omissions or changes / no information
Inclusion and exclusion criteria for participants	GAR-form; publication	Discrepancy-item 2	No discrepancies / criteria changed / no information
Primary, secondary, and other endpoints	Protocol; publication	Discrepancy-item 3	No discrepancies / primary endpoints added / primary endpoints omitted / primary endpoints changed / other additions, omissions or changes / no information
Planned and actual number of participants	GAR-form; publication	Discrepancy-item 4	No discrepancies / sample size smaller / sample size larger / no information
Methods for any additional analyses (e.g. subgroup)	Protocol; publication	Discrepancy-item 5	No discrepancies / analysis added / analyses omitted / analyses changed / no information
Intention to treat (ITT) or per protocol (PP) analysis	Protocol; publication	Discrepancy-item 6	No discrepancies / changed / no information
Sponsor acknowledgement	Publication	Discrepancy-item 7	Yes (specific sponsor) / no
Secondary publications: planned in protocol and mentioned in publication	Protocol; publication	Multiple publications	Planned / not planned and not mentioned / not planned and mentioned / no information

TABLE SA2.1.2 Data extraction form. GAR = General Assessment and Registration (continued)

Extract:	Source	Use	If categorical, options
Direction of publication conclusion	Publication	Direction of publication conclusion	Positive / negative
Prospective registration on clinicaltrials.gov	www. clinicaltrials.gov	Determinant	Yes / no

2

REFERENCES

1. World Medical Association. The Declaration of Helsinki. 2013.

2. Dixon JR, Jr. The International Conference on Harmonization Good Clinical Practice guideline. *Qual Assur.* Apr-Jun 1998;6(2):65-74.

3. Bardy AH. Bias in reporting clinical trials. *Br J Clin Pharmacol.* Aug 1998;46(2):147-150.

4. Benjamin DK, Jr., Smith PB, Murphy MD, et al. Peer-reviewed publication of clinical trials completed for pediatric exclusivity. *JAMA.* Sep 13 2006;296(10):1266-1273.

5. Blumle A, Antes G, Schumacher M, Just H, von Elm E. Clinical research projects at a German medical faculty: follow-up from ethical approval to publication and citation by others. *J Med Ethics.* Sep 2008;34(9):e20.

6. de Jong JP, Ter Riet G, Willems DL. Two prognostic indicators of the publication rate of clinical studies were available during ethical review. *Journal of clinical epidemiology.* Dec 2010;63(12):1342-1350.

7. Decullier E, Chan AW, Chapuis F. Inadequate dissemination of phase I trials: a retro-spective cohort study. *PLoS Med.* Feb 17 2009;6(2):e1000034.

8. Decullier E, Lheritier V, Chapuis F. Fate of biomedical research protocols and publica-tion bias in France: retrospective cohort study. *BMJ.* Jul 2 2005;331(7507):19.

9. Easterbrook PJ, Matthews DR. Fate of research studies. *J R Soc Med.* Feb 1992;85(2):71-76.

10. Hemminki E. Study of information submitted by drug companies to licensing authorities. *Br Med J.* Mar 22 1980;280(6217):833-836.

11. Hole OP, Nitter-Hauge S, Cederkvist HR, Winther FO. An analysis of the clinical devel-opment of drugs in Norway for the year 2000: the completion of research and publication of results. *Eur J Clin Pharmacol.* Mar 2009;65(3):315-318.

12. Jones CW, Handler L, Crowell KE, Keil LG, Weaver MA, Platts-Mills TF. Non-publication of large randomized clinical trials: cross sectional analysis. *BMJ.* 2013;347:f6104.

13. Kasenda B, von Elm E, You J, et al. Prevalence, characteristics, and publication of dis-continued randomized trials. *JAMA.* Mar 12 2014;311(10):1045-1051.

14. Lee K, Bacchetti P, Sim I. Publication of clinical trials supporting successful new drug applications: a literature analysis. *PLoS Med.* Sep 23 2008;5(9):e191.

15. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine--selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ.* May 31 2003;326(7400):1171-1173.

16. Pich J, Carne X, Arnaiz JA, Gomez B, Trilla A, Rodes J. Role of a research ethics commit-tee in follow-up and publication of results. *Lancet.* Mar 22 2003;361(9362):1015-1016.

17. Shamliyan T, Kane RL. Clinical research involving children: registration, completeness, and publication. *Pediatrics.* May 2012;129(5):e1291-1300.

18. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ.* Sep 13 1997;315(7109):640-645.

19. von Elm E, Rollin A, Blumle A, Huwiler K, Witschi M, Egger M. Publication and non-publication of clinical trials: longitudinal study of applications submitted to a research ethics committee. *Swiss Med Wkly.* Apr 5 2008;138(13-14):197-203.

20. Ramsey S, Scoggins J. Commentary: practicing on the tip of an information iceberg? Evidence of underpublication of registered clinical trials in oncology. *Oncologist.* Sep 2008;13(9):925-929.

21. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet.* Apr 13 1991;337(8746):867-872.

22. Cowley AJ, Skene A, Stainer K, Hampton JR. The effect of lorcainide on arrhythmias and survival in patients with acute myocardial infarction: an example of publication bias. *International journal of cardiology.* Jul 1 1993;40(2):161-166.

23. Nowbar AN, Mielewczik M, Karavassilis M, et al. Discrepancies in autologous bone marrow stem cell trials and enhancement of ejection fraction (DAMASCENE): weighted regression and meta-analysis. *BMJ.* 2014;348:g2688.

24. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA.* Jan 15 1992;267(3):374-378.

25. Every-Palmer S, Howick J. How evidence-based medicine is failing due to biased trials and selective publication. *J Eval Clin Pract.* May 12 2014.

26. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ.* 2010;340:c365.

27. Saini P, Loke YK, Gamble C, Altman DG, Williamson PR, Kirkham JJ. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ.* 2014;349:g6501.

28. Smyth RM, Kirkham JJ, Jacoby A, Altman DG, Gamble C, Williamson PR. Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists. *BMJ.* 2011;342:c7153.

29. *Institute of Medicine (IOM) Discussion framework for clinical trial data sharing: Guiding principles, elements, and activities.* Washington, DC: The National Academies Press;2014.

30. Ioannidis JP. How to make more published research true. *PLoS Med.* Oct 2014;11(10):e1001747.

31. Lee KP, Boyd EA, Holroyd-Leduc JM, Bacchetti P, Bero LA. Predictors of publication: characteristics of submitted manuscripts associated with acceptance at major biomedical journals. *Med J Aust.* Jun 19 2006;184(12):621-626.

32. Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA.* Jan 24-31 2001;285(4):437-443.

33. Kasenda B, Schandelmaier S, Sun X, et al. Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications. *BMJ.* 2014;349:g4539.

34. Dwan K, Altman DG, Clarke M, et al. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. *PLoS Med.* Jun 2014;11(6):e1001666.

35. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA.* May 26 2004;291(20):2457-2465.

36. Chan AW, Hrobjartsson A, Jorgensen KJ, Gotzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ.* 2008;337:a2299.

37. Hahn S, Williamson PR, Hutton JL. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. *J Eval Clin Pract.* Aug 2002;8(3):353-359.

2

38. Eyding D, Lelgemann M, Grouven U, et al. Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. *BMJ.* 2010;341:c4737.

39. McGauran N, Wieseler B, Kreis J, Schuler YB, Kolsch H, Kaiser T. Reporting bias in medical research - a narrative review. *Trials.* 2010;11:37.

40. Lexchin J. Those who have the gold make the evidence: how the pharmaceutical industry biases the outcomes of clinical trials of medications. *Science and engineering ethics.* Jun 2012;18(2):247-261.

41. Getz, K.A. Getting to the Heart of Public Distrust. http://www.appliedclinicaltrialsonline. com/appliedclinicaltrials/Drug+Development/Getting-to-the-Heart-of-Public-Distrust/ ArticleStandard/Article/detail/546111, accessed 12-9-2014.

42. ICMJE. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals (Update December 2013). Accessed 25-08-2014. www. IMCJE.org.

43. Chan AW, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med.* Feb 5 2013;158(3):200-207.

44. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340:c332.

45. Prayle AP, Hurley MN, Smyth AR. Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study. *BMJ.* 2012;344:d7373.

46. Huic M, Marusic M, Marusic A. Completeness and changes in registered data and reporting bias of randomized controlled trials in ICMJE journals after trial registration policy. *PLoS One.* 2011;6(9):e25258.

47. Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS One.* 2013;8(7):e66844.

48. http://www.ccmo.nl/en/accredited-mrecs [accessed on: 20-11-2014].

49. World Health Organization. International Classfication of Diseases Version 10. http:// apps.who.int/classifications/icd10/browse/2015/en; lastly accessed on 12 January 2015.

50. Communication from the Commission — Detailed guidance on the request to the competent authorities for authorisation of a clinical trial on a medicinal product for human use, the notification of substantial amendments and the declaration of the end of the trial: article 43 (CT-1). Official Journal of the European Union2010.

51. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet.* Jul 4 2009;374(9683):86-89.

52. Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet.* Jan 18 2014;383(9913):267-276.

53. Chan AW, Song F, Vickers A, et al. Increasing value and reducing waste: addressing inaccessible research. *Lancet.* Jan 18 2014;383(9913):257-266.

Appendix 2.2

Questionnaire to the investigator

2

Questionnaires were sent to the principal investigators of all 622 clinical drug trials that were reviewed by a Dutch Institutional Review Board in 2007. Four different questionnaires were used, based on our search results.

1. We found no publication and no data on whether the trial was completed as planned or discontinued before planned end of recruitment and/or follow-up
2. We found a publication and no data on whether the trial was completed as planned or discontinued before planned end of recruitment and/or follow-up
3. We found no publication and data on whether the trial was completed as planned or discontinued before planned end of recruitment and/or follow-up
4. We found at least one publication and data on whether the trial was completed as planned or discontinued before planned end of recruitment and/or follow-up.

In this appendix, the template of questionnaire #1 is included. The other questionnaires were copies of questionnaire 1, except for the abovementioned adaptations.

QUESTIONNAIRE 1

UPPER survey 2.0 - Questionnaire of research project 'Better Use of Files'

Questionnaire of research project 'Better Use of Files'

This is the online questionnaire of the research project 'Better Use of Files, of which you were informed by email

There are 16 questions in this survey

Trial identification

1 [1]

What is the NL-number of the trial?

Please fill out the NL-number exactly as it is noted in the email, including the 2 dots. Example: NL12345.678.90

Investigators who received emails for more than one trial, please complete a separate survey for each trial *

Please check the format of your answer.

Please write your answer here:

2 [8]What was the date that inclusion started? *

Please write your answer(s) here:

In the Netherlands:

In case of an international multicenter trial: for the

whole trial:

3 [4]Has the trial prospectively been registered at a public register, e.g. www.clinicaltrials.gov or ISRCTN? By prospectively we mean before recruitment of the first participant. *

Please choose **only one** of the following:

○ Yes

UPPER survey 2.0 - Questionnaire of research project 'Better Use of Files'

○ No

---

4 [6]Note the trial identification number for the register at issue (e.g. NCT12345678) *

Only answer this question if the following conditions are met:
Answer was 'Yes' at question '3 [4]' (Has the trial prospectively been registered at a public register, e.g. www.clinicaltrials.gov or ISRCTN? By prospectively we mean before recruitment of the first participant.)

Please write your answer here:

Start of the trial

These questions address whether the trial was started after approval of the medical research ethics committee

---

5 [1]Are there any participants included in this trial? *

Please choose **only one** of the following:

○ Yes, also in the Netherlands
○ Not in the Netherlands, but in other countries
○ No, the trial has not started at all

---

6 [2]Why has inclusion for the trial not started in the Netherlands? *

Only answer this question if the following conditions are met:
Answer was 'Not in the Netherlands, but in other countries' at question '5 [1]' (Are there any participants included in this trial?)

Please choose all that apply and provide a comment:

☐ There were already sufficient participants included
in other countries

☐ Other:

In the window on the right you can give further explanation, if desired

---

7 [5]Why has the trial not started at all? *

Only answer this question if the following conditions are met:
Answer was 'No, the trial has not started at all' at question '5 [1]' (Are there any participants included in this trial?)

Please write your answer here:

UPPER survey 2.0 - Questionnaire of research project 'Better Use of Files'

Progress of the trial

These questions address the progress of the trial

8 [1]How many participants were included in the trial?

Only answer this question if the following conditions are met:
Answer was 'Not in the Netherlands, but in other countries' *or* 'Yes, also in the Netherlands' at question '5 [1]' (Are there any participants included in this trial?)

Please write your answer(s) here:

UPPER survey 2.0 - Questionnaire of research project 'Better Use of Files'

Number of participants included in the Netherlands:	
If the trial is an international multicenter trial: number of participants included in whole trial:	

9 [3]How was the trial completed? *

Only answer this question if the following conditions are met:
Answer was 'Yes, also in the Netherlands' *or* 'Not in the Netherlands, but in other countries' at question '5 [1]' (Are there any participants included in this trial?)

Please choose **only one** of the following:

○ Completed as planned
○ Preliminary terminated

10 [4]What was the reason for the preliminary termination of the trial? *

Only answer this question if the following conditions are met:
Answer was 'Preliminary terminated' at question '9 [3]' (How was the trial completed?) *and* Answer was 'Yes, also in the Netherlands' *or* 'Not in the Netherlands, but in other countries' at question '5 [1]' (Are there any participants included in this trial?)

Please choose all that apply and provide a comment:

☐ Insufficient participants

☐ Adverse events

☐ Interim analysis shows superiority

☐ Interim analysis shows futility

☐ Stopped on advice of the Data Monitoring Board

☐ Other:

Multiple answers can be selected and explanation may be provided in the boxes on the right

11 [5]What was the date of the last visit of the last patient in the trial? Please input the date in the following format: dd-mm-yyyy, e.g. 01-01-2010 *

Only answer this question if the following conditions are met:
Answer was 'Not in the Netherlands, but in other countries' *or* 'Yes, also in the Netherlands' at question '5 [1]' (Are there any participants included in this trial?)

Please write your answer(s) here:

In the Netherlands:	
If the trial is an international multicenter trial: for the whole trial:	

UPPER survey 2.0 - Questionnaire of research project 'Better Use of Files'

For example: 01-01-2010

Publication of the trial

The final set of questions addresses the publication of the trial. By publication we mean publishing the methods and results of the trial in a peer-reviewed journal

12 [1]We hebben in Medline, Embase en Google Scholar geen publicatie van de studie kunnen vinden. Klopt het dat de studie (nog) niet als artikel is gepubliceerd? *

Only answer this question if the following conditions are met:
Answer was 'Yes, also in the Netherlands' *or* 'Not in the Netherlands, but in other countries' at question '5 [1]' (Are there any participants included in this trial?)

Please choose **only one** of the following:

○ No, the trial has been published in a peer-reviewed journal
○ Yes, the trial has indeed not been published

13 [2]

Please note the title(s), journal name(s), journal edition(s), and publication date(s) of the publication(s) *

Only answer this question if the following conditions are met:
Answer was 'No, the trial has been published in a peer-reviewed journal' at question '12 [1]' (We hebben in Medline, Embase en Google Scholar geen publicatie van de studie kunnen vinden. Klopt het dat de studie (nog) niet als artikel is gepubliceerd?) *and* Answer was 'Yes, also in the Netherlands' *or* 'Not in the Netherlands, but in other countries' at question '5 [1]' (Are there any participants included in this trial?)

Please write your answer here:

UPPER survey 2.0 - Questionnaire of research project 'Better Use of Files'

---

14 [3]

Why has the trial not been published? *

Only answer this question if the following conditions are met:
Answer was 'Yes, the trial has indeed not been published' at question '12 [1]' (We hebben in Medline, Embase en Google Scholar geen publicatie van de studie kunnen vinden. Klopt het dat de studie (nog) niet als artikel is gepubliceerd?) *and* Answer was 'Not in the Netherlands, but in other countries' *or* 'Yes, also in the Netherlands' at question '5 [1]' (Are there any participants included in this trial?)

Please choose all that apply and provide a comment:

☐ Andere prioriteiten

☐ Resultaten niet statistisch significant

☐ Resultaten niet klinisch relevant

☐ Manuscript is afgewezen door tijdschrift(en)

☐ Andere reden(en):

---

15 [4]Are the results of the trial reported in a different way than in a peer reviewed journal?

Only answer this question if the following conditions are met:
Answer was 'Yes, also in the Netherlands' *or* 'Not in the Netherlands, but in other countries' at question '5 [1]' (Are there any participants included in this trial?)

UPPER survey 2.0 - Questionnaire of research project 'Better Use of Files'

Please choose **only one** of the following:

○ Yes (Please specify where, e.g. results reported on clinicaltrials.gov + registration number)

○ No
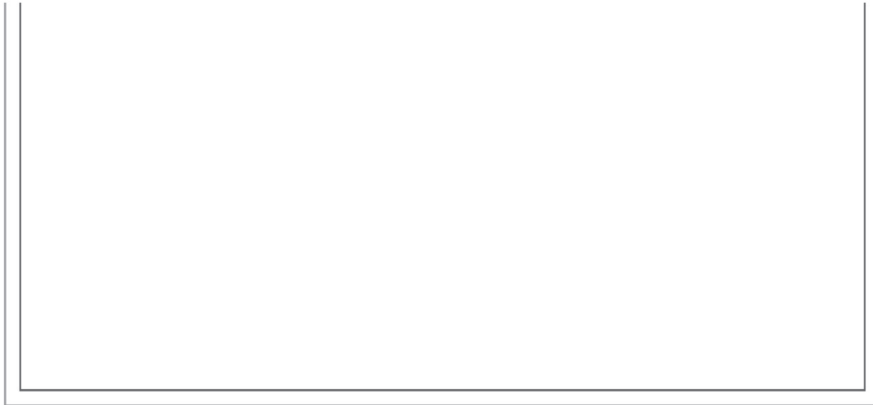
Make a comment on your choice here:

Einde vragenlijst

16 [3]Thank you for completing the questionnaire. You can add any additional comments here. Please do not forget to click on the button 'Submit' at the bottom of this page.

Please write your answer here:

2

UPPER survey 2.0 - Questionnaire of research project 'Better Use of Files'

Many thanks for your collaboration! Your answers have been saved. Please do not fill out the questionnaire again.

In case of unforeseen errors or system malfunction, please contact Sander van den Bogert: s.van.den.bogert@ccmo.nl

Submit your survey.
Thank you for completing this survey.

Appendix 2.3

Trends in the clinical drug trial
landscape in the Netherlands

2

INTRODUCTION

This thesis includes several studies of the Dutch clinical drug trial landscape of 2007. All drug trials reviewed by an institutional review board (IRB) in 2007 were studied. As the clinical trial landscape is dynamic and may evolve over time, characteristics of clinical drug trials in 2007 may look different compared to trials reviewed by IRBs in more recent years. This potential change in the outlook of the trial landscape may impact the interpretation of the results and recommendations for research and practice. Therefore, in this appendix to the chapter of the 2007 cohort, we aimed to investigate trends in the characteristics of clinical drug trials in the Netherlands after 2007, and to examine if changes may influence the findings of the studies of the 2007 cohort described in chapter 2.

METHODS

From ToetsingOnline we extracted all clinical drug trials reviewed by all accredited Dutch IRBs from 1 January 2007 until 31 December 2015. Extracted trial characteristics included IRB decision; sponsor; phase, contract research organization (CRO) as applicant; number and location of centers; therapeutic effect expected; type of trial; approval status of the drugs involved in the trial; drug type; participant category; and the disease area.

First, the absolute number of reviewed trials per year was assessed using a marked line graph. Then, the proportional distribution of characteristics was depicted per year of IRB approval (2007, 2008,..., 2015) using stacked column graphs. The trends were assessed by visual inspection of their graphical representation. Graphs were created using Microsoft Excel.

RESULTS

The number of clinical drug trials that were reviewed by all accredited Dutch IRBs between 2007 and 2015 varied between 635 (2008) and 528 (2010), as shown in figure A2.2.1. After the 2010 drop, from 625 to 528 trials, the numbers seem to recover towards the level of 2007-2009.
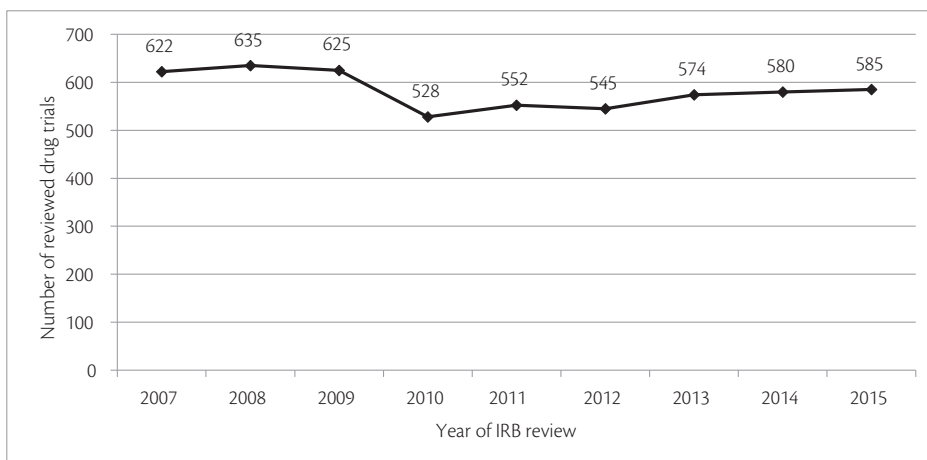
FIGURE A2.3.1 Number of clinical drug trials reviewed by all Dutch IRBs by year.

The sponsoring of the trials remained stable over time (figure A2.2.2). There was a slight drop of investigator-initiated trials without industry (co-)funding in 2015, but it is unknown whether this is structural. Regarding the disease areas (figure A2.2.3), the share of each area in the overall landscape varied per year. The disease areas that seemed to show a consistent increase included oncology (from 18.8% in 2007 to 24.3% in 2015) and hematological and immunological diseases (from 6.1% in 2007 to 10.3% in 2015). Disease areas that showed an overall consistent decrease included cardiovascular diseases (from 10.9% in 2007 to 6.3% in 2015) and mental/behavioral diseases (from 7.2% in 2007 to 4.4% in 2015).

No changes over time were observed in the characteristics IRB decision; phase; CRO as applicant; therapeutic effect expected; study type; approval status of drugs involved in the trial; drug type; and participant category (results not shown).
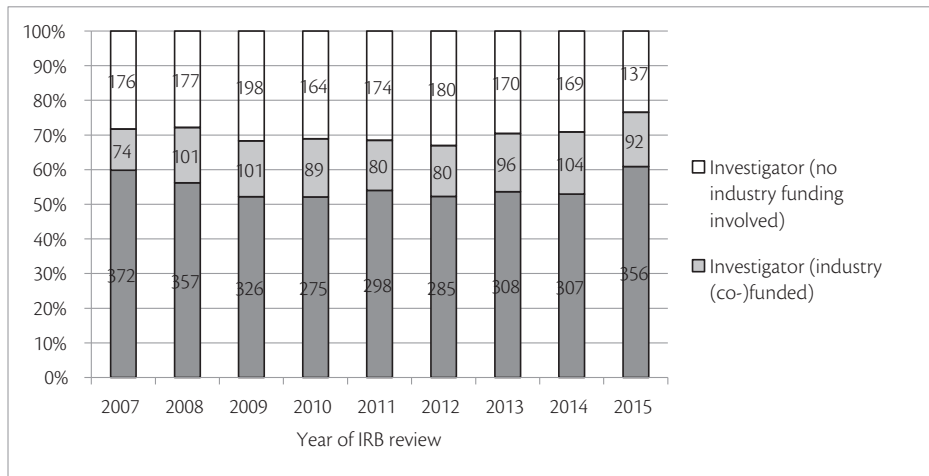
FIGURE A2.3.2 Distribution of the sponsor type of the clinical drug trials by year of IRB-review. The numbers in the bars are the number of trials in that year with the corresponding sponsor type.

DISCUSSION

Overall, the clinical drug trial landscape in the Netherlands has remained stable between 2007 and 2015. The drop in 2010 could have been the result of the financial crisis that hit Europe around that time – as a consequence, companies and investigators might have had less financial space to start new projects [1]. Furthermore, outsourcing trials to cheaper countries may be particularly popular in times of crisis. Since 2010, the number of trials gradually recovered, approximating its pre-2010 level in 2015.

The rise of oncology in clinical research was expected and is in line with other analyses. The increased trial activity in hematological and immunological diseases likely also reflects the development of new pharmacological immunotherapies (e.g. peptides, antibodies). The decrease of research in cardiovascular diseases is consistent with negative forecasts of R&D productivity for these areas [2 3].

The results of this analysis only provide insights in trends of the analyzed characteristics, and not about the outcomes that were investigated in the studies of this chapter. We also cannot claim based on these results that the clinical trial landscape has not changed. Reality is closer to the opposite regarding the average number of countries per trial, recruitment patterns and endpoints [4]. The aims and scope of this analysis was limited to the investigation of the representativeness of the 2007 cohort in the recent history of clinical drug trials in the Netherlands, i.e. to examine the possibility that 2007 was an outlier regarding the investigated characteristics.

To conclude, our analysis of the clinical drug trial landscape of the Netherlands in the years after 2007 suggests that the landscape has remained stable. The distributions of characteristics of the later years show that the findings of the studies described in chapter 2 of this thesis are based on a representative cohort.

FIGURE A2.3.3 Distribution of the disease areas of the clinical drug trials by year of IRB-review. The numbers in the bars are the number of trials in that year in the corresponding disease area.

REFERENCES

1. Getz KA. Is Clinical Research Recession Proof? Applied Clinical Trials 2009

2. Smietana K, Ekstrom L, Jeffery B, Moller M. Improving R&D productivity. Nat Rev Drug Discov 2015;**14**(7):455-6 doi: 10.1038/nrd4650[published Online First: Epub Date]|.

3. Kesselheim AS, Hwang TJ, Franklin JM. Two decades of new drug development for central nervous system disorders. Nat Rev Drug Discov 2015;**14**(12):815-6 doi: 10.1038/nrd4793[published Online First: Epub Date]|.

4. Rosenblatt M. The Large Pharmaceutical Company Perspective. N Engl J Med 2017;**376**(1):52-60 doi: 10.1056/NEJMra1510069[published Online First: Epub Date]|.

2

Chapter 3

Determinants of selective reporting.
A taxonomy based on content analysis
of a random selection of the literature

Authors
Jenny T. van der Steen; Cornelis A. van den Bogert; Mirjam C. van Soest-
Poortvliet; Soulmaz Fazeli Farsani; René H.J. Otten; Gerben ter Riet; Lex
M. Bouter

ABSTRACT

Objectives

Selective reporting is wasteful, leads to bias in the published record and harms the credibility of science. Studies on potential determinants of selective reporting currently lack a shared taxonomy and a causal framework. To develop a taxonomy of determinants of selective reporting in science.

Methods

Using search terms for bias and selection combined with terms for reporting and publication, we systematically searched the PubMed, Embase, PsycINFO and Web of Science databases up to January 8, 2015. Of the articles identified, we screened a 25 percent random selection. From eligible articles, we extracted phrases that mentioned determinants of selective reporting, which we used to create meaningful categories. We stopped when no new categories had emerged in the most recently analyzed articles (saturation).

Results

Saturation was reached after analyzing 64 articles. We identified 497 putative determinants, of which 145 (29%) were supported by empirical findings. The determinants represented 12 categories (leaving 3% unspecified): focus on preferred findings (36%), poor or overly flexible research design (22%), high-risk area and its development (8%), dependence upon sponsors (8%), prejudice (7%), lack of resources including time (3%), doubts about reporting being worth the effort (3%), limitations in reporting and editorial practices (3%), academic publication system hurdles (3%), unfavorable geographical and regulatory environment (2%), relationship and collaboration issues (2%), and potential harm (0.4%).

Conclusions

We designed a taxonomy of putative determinants of selective reporting consisting of 12 categories. These categories feature in a novel theory about causes of selection bias, in which motives (focus, prejudice) and means (design) serve as necessary causes. Our theory provides a basis for hypothesis testing in future research. The taxonomy and theory might also guide policies to prevent selective reporting.

INTRODUCTION

Complete, accurate and timely reporting of all (study protocol-specified) outcomes is essential for syntheses of research to be valid and as precise as possible[1,2]. Complete or unselective reporting refers to both (unselective) publication of all results of a study as well as unselective or complete reporting within publications on all planned outcomes[3]. In other words, all planned outcomes should be reported on within a reasonable period (and the exploratory nature of analyses with any unplanned outcomes should be disclosed).

Selective reporting leads to bias if specific results remain unpublished because the decision to report depends on the nature of the results (e.g., direction or magnitude of the target association). Reporting bias is an important threat to the validity of systematic reviews which clinicians, researchers, policy makers and citizens, rely on. Therefore, reporting bias is wasteful, distorts the aggregate body of scientific evidence, threatens the credibility of science, but it may also result in suboptimal treatment or even in avoidable harm to e.g., patients' health. Therefore, in addition to validity and efficiency reasons, there is an ethical imperative of reporting all results including those of clinical trials [4].

Around half of planned outcomes of clinical trials are not reported, and a third to half of registered clinical trials remains unpublished[5-8]. There are numerous reports that suggest that selective publication is a major problem in the clinical domain, but it is also pervasive in basic and translational research [9,10] and in the social sciences [11]. Selective reporting occurs in various types of studies other types of studies and across various designs, e.g., trials with psycho-educational interventions, and quantitative observational and qualitative studies[8,12,13].

So-called "protocol-to-publication" and similar studies mention selective reporting of statistically significant results[8,14,15]. More generally–also considering, for example, equivalence trials and studies into adverse effects, where authors often prefer non-significance–"null" findings may be preferably reported. As a consequence, findings preferred by key stakeholders in the research project at issue are likely overrepresented in the scientific literature[16].

Research has suggested that financial conflicts of interest may cause selective reporting (e.g., when studies are sponsored by industry[6,8,17,18]), but non-financial conflicts of interest probably play an important role too. Causes of publication bias, and of reporting bias more generally, may relate to decisions taken by researchers and sponsors, and also decisions by editors[5,6,19,20]. Some have argued that it is the human nature to search for positive messages, which suggests that basically, all scientists are at risk[2,21]. However, certain persons or environments may be at increased risk of selective reporting such as junior researchers[22] and

scientists in more competitive academic environments [23]. Despite numerous studies on selective reporting, there is no accepted taxonomy of its determinants and no explicit causal framework.

We aimed at developing a taxonomy of putative determinants of selective reporting. We therefore addressed the questions what are possible determinants of selective reporting, and how putative determinants of selective reporting might be grouped best by its content.

METHODS

Design

To develop a taxonomy of putative determinants of selective reporting, we combined principles of systematic reviews[24] with those of inductive qualitative content analysis[25,26]. Before analyzing full-texts, we piloted search strategies, abstract and full-text eligibility criteria and reviewed procedures. We developed the study protocol based on the preferred reporting items for systematic review and meta-analysis *protocols* (PRISMA-*P*) guidelines[27,28] as far as applicable. In this article, we report applicable items from the PRISMA[29] guidelines for systematic review and applicable items from guidelines for the reporting of qualitative studies[30,31].

Eligibility criteria

We included articles that examined or suggested determinants of selective reporting. We sought for articles from any academic discipline reporting on studies employing any type of design based on empirical data, such as intervention studies (with any type of comparator), and observational studies. Additionally, to cover hypotheses on what drives selective reporting, we included non-empirical articles such as editorials presenting opinion, theoretical considerations and anecdotal evidence. To minimize duplication of determinants extracted from articles that were also selected for inclusion in reviews, we excluded review articles. The outcome, selective reporting, comprised non-publication and selective reporting within publications. Our review also covered the possible consequences of selective reporting, including publication bias and other types of reporting bias.

Information sources

We searched PubMed, Embase, PsycINFO and Web of Science to cover a wide range of academic disciplines from inception to January 8th 2015. We limited the search to the English, French, German and Dutch languages.

Search strategy

The search strategy focused on selective reporting to avoid limitation to any preconceived determinants. We used terms for bias and selection combined with terms for reporting and publication, and we pilot tested the search strategy.

For reasons of feasibility of identifying determinants with qualitative content analysis, we retained a 25% random sample of the 918 hits after de-duplication, from all databases searched from inception to 8 January 2015 (figure 3.1). We used SPSS version 22's random allocator function "random sample of cases" to randomly select a quarter comprising 230 hits.

Selection of articles

Titles and abstracts were screened independently against the inclusion criteria by reviewers (JTvdS, CAvdB and MCvS-P). If there was no abstract, we reviewed keywords. We used the first ten percent of abstracts (23) to test a shared understanding of the inclusion criteria and we discussed discrepant interpretations. Of the other abstracts, we calculated inter-rater agreement (percentage) of decisions on retrieval as full-text.

Data collection process

Full-text data extractions were performed by pairs of researchers (JTvdS, CAvdB, MCvS-P, SFF, and GtR). We evaluated for eligibility all full-text articles and we calculated the inter-rater agreement also for this step. JTvdS, CAvdB and MCvS-P piloted a standardized full-text data extraction spreadsheet using three empirical and non-empirical articles not included in the random sample. Initially, the reviewers extracted data independently, but when determinants were identified consistently after having analyzed several articles, one reviewer extracted data with verification by another.

Data items

At the level of the article, we abstracted year of publication, academic discipline and study design. More than one determinant per article was possible. We assessed if evidence about association of a determinant with the outcome was empirical, indicated an actor (stakeholder), referred to a cause or could possibly be interpreted as a cause (we thus explicitly assessed the degree of interpretation[32]).

We assessed the type and scope of selective reporting (whether limited to a single medium, e.g., a specific journal) and the nature of the association between a putative determinant and selective reporting (if there was any association, the direction, and strength of association if applicable).
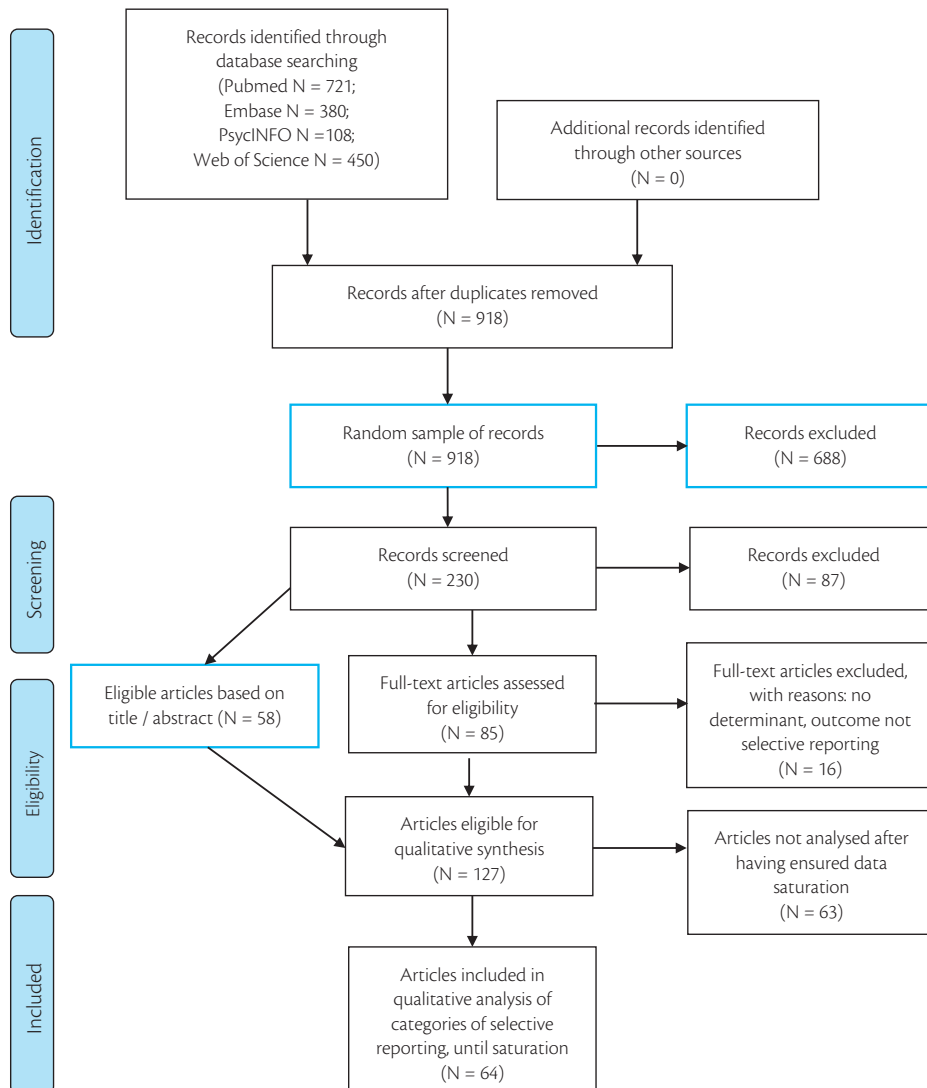
FIGURE 3.1 PRISMA flow diagram of identified and analyzed articles.

Analyses

The sample of articles and determinants are summarized using percentages. We used qualitative iterative and inductive content analysis to group determinants by content and form categories of determinants[26]. In particular, the reviewers (pair) extracted putative determinants in the context of the article, and the team subsequently discussed interpretations and categorization. In the content analysis, we coded putative determinants and subsumed them in meaningful categories without imposing any prior model. We also considered determinants that were tested but authors found them unrelated to selective reporting. A single dataset with putative determinants was created after discussions on differences of interpretation.

Next, two researchers (JTvdS and CAvdB) independently categorized all determinants into higher-level groupings. This was a non-linear, iterative process as we classified batches of about 50 determinants followed by discussing all classifications of each batch before moving on to the next batch. We avoided overlap in the categories[32] by adding descriptions (to serve as definitions) that we developed from the iterative classification of content. In reaching consensus about the categorization, we often went back to previous classification work to adapt categories, or to the full-texts to ensure we understood the context. Thus, we constructed a structured list of categories of more specific determinants. An initial classification, as were any unresolved issues about classifications, was discussed with a third researcher (GtR) to achieve full consensus. Following the principles of content analyses[26], we counted the number of determinants per category for descriptive purposes of which categories were and which were less prominent in the literature.

Saturation

We analyzed content of articles randomly selected from different decades (< 1980, 1980s, 1990s, 2000-2009, 2010-2015). We concluded the analyses when saturation was reached. Saturated data ensure replicability in categories derived from content analyses, which in turn verifies and ensures comprehensiveness[32]. Saturation of categories was determined in two ways. First, during the process of analyzing the batches of determinants, we assessed saturation prospectively, during the content analyses, in the usual way for qualitative analyses (i.e., no new categories emerged in the last analyzed articles). After having analyzed about 50 articles, we deemed the newly classified determinants not to be essentially different from those already classified (we could fit them in the categories we had developed), and we assumed saturation. This was confirmed after having analyzed one more batch of determinants representing 10% of articles (13 of 127). Second, retrospectively, after having analyzed the articles, we verified quantitatively how many articles had been analyzed when the first determinant of each category emerged. The first

3

determinant in each category usually emerged after analysis of only a few articles (figure 3.2). After having extracted 120 determinants (of 497, 24%) from 14 articles, all categories comprised one or more determinants.
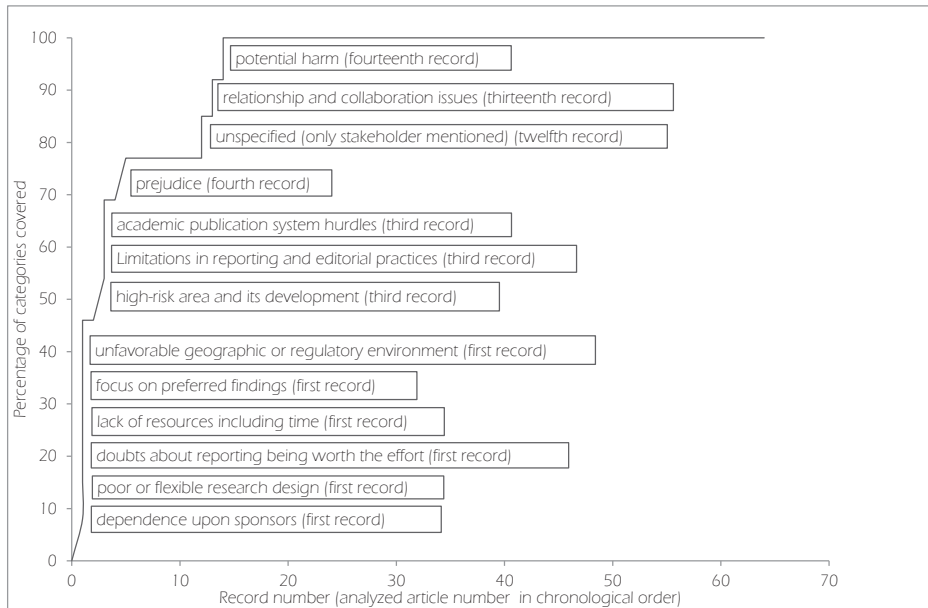

FIGURE 3.2 Saturation graph. The horizontal axis displays the articles (records) in chronological order of analysis. The vertical axis displays the percentage of the 12 (plus 1 unspecified) determinant categories containing at least one determinant. The labels describe which determinant category appeared for the first time in which record.

Subgroup analyses

We assessed the extent to which various determinants were based on empirical studies or (solely) on opinion. Further, we calculated the proportion of articles that (quantitatively or qualitatively) reporting a non-significant or no, or an unexpected direction of an association between a determinant and selective reporting.

RESULTS

We identified 918 unique records, and we included 64 records (articles) in the final analysis (figure 3.2). The inter-rater agreement about initial and independent assessment of the need to retrieve the full-text as assessed for 90% of the records was 72% (150/207; agreement about 143 records, doubt about 7, no agreement about 57). The initial agreement about eligibility of full-text assessment was 78% (66/85). All initial disagreements were resolved through discussion.

TABLE 3.1 Characteristics of the 64 analyzed articles.

Characteristic	N (%)
Year of publication	
< 1980	1 (2%)
1980s	2 (3%)
1990s	10 (16%)
2000-2009	28 (44%)
2010-2015	23 (36%)
Academic discipline	
Clinical medicine	48 (75%)
Biomedicine / Life sciences	6 (9%)
Humanities	7 (11%)
No specific discipline	3 (5%)
Type of study / study design	
Non-empirical (reflective / theoretical)	31 (48%)
Observational quantitative, longitudinal[*]	17 (27%)
Observational quantitative, cross-sectional[*]	6 (9%)
Case description / anecdotal	3 (5%)
Simulation / modelling[†]	3 (5%)
Randomized trial[‡]	1 (2%)
Qualitative	1 (2%)
Mixed methods (integrated)	1 (2%)
Review of reviews	1 (2%)

[*] Observational quantitative studies included: comparisons of publications (n=5), comparison of registry records with publications (4), of protocols with publications (4), of submitted with accepted papers (4), of abstracts with publications (4), of public funder database with publications (1) and of industry database with Medline records (1). [†] Mathematical simulations of reporting bias, subjective decision-making in peer-review, and the selection process in publication bias, whether purely hypothetical or with use of empirical data. [‡] The RCT assessed the effect of blinded peer-review on reviewers' and editors' decisions about manuscript acceptance [33]. The determinant was prejudice in the peer-review process, and the outcome was non-publication considering if the editor's decision dictates whether the manuscript is being published

Of the 64 articles analyzed, 48 (75%) concerned biomedicine / life sciences, and 51 (80%) were published in 2000 or later (table 3.1). Half of the articles (32) were non-empirical. The empirical studies were mostly observational and quantitative; we found one RCT[33].

We extracted 497 determinants from the 64 articles (median 6; range 1 (9 articles) to 22 (3 articles; table 3.2). Twenty-nine percent (145 determinants) concerned empirical evidence about associations with selective reporting. If an actor (stakeholder) was mentioned (41%, 204 determinants), it was the investigator in about half of cases (110 determinants). In 79% of cases, an association with

selective reporting was found or postulated (all in the hypothesized or expected direction).

TABLE 3.2 Characteristics of the 497 extracted determinants, outcomes and their associations.

Characteristic	N (%)
Evidence about association of determinant with outcome	
Empirical	145 (29%)
Non-empirical (e.g. from viewpoint, or opinion in discussion section, or inference from the literature or theoretical study)	352 (71%)
Actor (stakeholder) involved	204 (41%)
Investigators or authors	110 (22%)
Editors or journals	57 (11%)
Reviewers	18 (4%)
Sponsors or industry	17 (3%)
Government	1 (0.2%)
Analyst	1 (0.2%)
No actor mentioned	293 (59%)
Interpretation of association in terms of possible causal pathways[*]	
Describes a cause	111 (22%)
Allows for a single and clear interpretation of cause	98 (20%)
Unclear cause or multiple causal interpretations are possible	288 (58%)
Type of selective reporting outcome	
Non-publication	292 (59%)
Selective publication in general	78 (16%)
Selective reporting within publication	70 (14%)
Reporting bias	53 (11%)
Other (delayed publication, which risks e.g. late uptake in reviews)	4 (1%)
Scope of selective reporting outcome	
Within a single medium (journal or conference)	66 (13%)
General	431 (87%)
Reported association between determinant and outcome	
Present	393 (79%)
No association	104 (21%)

[*] Examples: "The company often owns the study database and controls decisions about publication and release of data" (describes a cause); authors reported "lack of time" as a reason (allows for a single and clear interpretation of cause– a cause is implicated (lack of time), but that cause itself begs a more detailed explanation (how does the reporting compete with other duties and why?)); sample size (unclear cause or multiple causal interpretations are possible– such as with larger sample size more power, more collaborators, more rigorous design, more quality checks etc.).

Table 3.3 lists the 12 categories that emerged from coding the extracted determinants, along with descriptions and examples. *Focus on preferred findings* was the largest category (180 determinants, 36%), which included, for example, significance chasing. This concerned empirical data in 17% (30/180) of cases, and in 93% (168/180) of cases, the original authors postulated that a focus on preferred findings was positively associated with selective reporting. By contrast, the second largest category (109 determinants, 22%), *poor or flexible research design*, was based on empirical findings in half of the cases, and the original authors mentioned a positive association in 57% of cases. The other 10 categories occurred less frequently (8% or less) yet represented distinct concepts. References to the 64 analyzed articles are provided per category, as a supplement (Supplementary results).

TABLE 3.3 Taxonomy of determinants (n=497) resulting from the inductive qualitative content analyses.

Determinant classification, category	Description	Examples	% (n in full sample)	% empirical result[*] (N in category /-per row)	% any relationship[†] (N in category / per row)
1. Focus on preferred findings	Focused on finding results that match preferences, mostly statistically significant or positive findings in other ways, wishful thinking and acting	Significance chasing, significant results, effect size, suppressing publication of unfavorable results, not being intrigued by null findings	36% (180/497)	17% (30/180)	93% (168/180)
2. Poor or flexible research design	Attributes of study design relating to power and level of evidence providing leeway in performing studies and its interpretation	Not a controlled or blinded study, study protocol unavailable, small sample size	22% (109/497)	50% (54/109)	57% (62/109)
3. High-risk area and its development	Area of research or discipline or specialty including historical development and competitiveness, the currently dominant paradigms and designs, and career opportunities	Ideological biases in a research field, area with much epidemiological research versus "hard sciences", humanities, experimental analytic methods, hot fields, publication pressure in the specific field	8% (39/497)	31% (12/39)	72% (28/39)

TABLE 3.3 Taxonomy of determinants (n=497) resulting from the inductive qualitative content analyses. (continued)

Determinant classification, category	Description	Examples	% (n in full sample)	% empirical result[*] (N in category /-per row)	% any relationship[†] (N in category / per row)
4. Dependence upon sponsors	Financial conflict of interest relating to lack of academic freedom	Requirements and influence of funding source with financial interests in the results	8% (38/497)	34% (13/38)	82% (31/38)
5. Prejudice	Belief that may be unfounded, whether aware or not	Prior belief about efficacy of treatment, author reputation or gender in the phase of review	7% (33/497)	24% (8/33)	82 (27/33)
6. Lack of resources including time	Insufficient manpower or finances	Excessive workload causes lack of time, or life events cause lack of personnel	3% (17/497)	18% (3/17)	100% (17/17)
7. Doubts about reporting being worth the effort	Weighing investment of time and means versus likelihood of gain through publication	Anticipating disappointment of yet another rejection versus low chances of acceptance of a manuscript, belief that findings are not worth the trouble	3% (16/497)	6% (1/16)	100% (16/16)
8. Limitations in reporting and editorial practices	Constraints and barriers to the practice of reporting relevant detail	Journal space restrictions, author writing skills	3 (14/397)	71% (10/14)	50% (7/14)
9. Academic publication system hurdles	Various hurdles to full reporting related to submission and processing of manuscripts (other than reporting) including those that represent an intellectual conflict of interest	Solicited manuscripts, authors indicating non-preferred reviewers, editor's rejection rate	3% (14/497)	36% (5/14)	57% (8/14)
10. Unfavorable geographical or regulatory environment[‡]	Geographical or regulatory environment that affects how research is being performed	Continent North America, Europe or Asia; few international collaborations; no governmental regulation of commercially sponsored research.	2% (12/497)	67% (8/12)	75% (9/12)

TABLE 3.3 Taxonomy of determinants (n=497) resulting from the inductive qualitative content analyses. (continued)

Determinant classification, category	Description	Examples	% (n in full sample)	% empirical result[*] (N in category /-per row)	% any relationship[†] (N in category / per row)
11. Relationship and collaboration issues	Intellectual conflict of interest between reporting and maintaining good relationships	Disagreements among co-authors and between authors and sponsors, sponsors select investigators who are biased towards the sponsor's position	2% (8/497)	13% (1/8)	100% (8/8)
12. Potential harm	Publishing data can harm individuals	Risk of bioterrorism, or confidentiality restriction	0.4% (2/497)	0% (0/2)	100% (2/2)
(13) Not specified	Referring to a stakeholder only	Selective publication not caused by editors	3% (15/497)	0% (0/15)	67% (10/15)

[footnotes to table 3.3] *Empirical result as described in table 3.2, first row. †Any relationship, and in the expected direction if any relationship was being hypothesized, versus no relationship. None of the hypothesized relationships in empirical result was found to be in the opposite direction. ‡ We phrased this category in terms of unfavorable because the work from which we abstracted the determinants for in particular this category often did not provide background or hypotheses of direction, probably being data driven rather than theory driven. For example, manuscripts from the US versus all other countries were tested and there were very few manuscripts from other countries.

The descriptions clarified boundaries between categories that were conceptually close, such as *high-risk area and its development*, and *unfavorable geographical and regulatory environment*. Both categories represented a wider environment than the research team or institution.

The categories of *unfavorable geographical and regulatory environment* and *academic publication system hurdles* were distinct as these included determinants not clearly referring to a possible hypothesis regarding a mechanism or cause of selective reporting (an example is provided as a footnote to table 3.3). By including these categories, we were able to classify all determinants, except for 15 that only mentioned a stakeholder (actor) as the source of selective reporting (or the denied source, in case of which we recorded no association between determinant and the outcome of selective reporting).

We found six described instances of interaction between determinants (effect modification). We counted these as classified with the main determinant only. The interactions all clearly described causes. For example, "Outcomes could be deemed post hoc to have little clinical relevance if they fail to show significant findings and may thus be omitted when accommodating space limitations"[34]. In

this case, the interaction between *focus on preferred findings* and *limitations in reporting and editorial practices* was classified under the first. By contrast, we classified another interaction, between a *focus on preferred findings* and *high-risk area and its development* under the latter as the primary category*: "*Early in the history of a research domain results in either direction are important news but that later, when the preponderance of evidence has supported one direction, significant reversals are often more important news than further replications"[35].

DISCUSSION

We developed a taxonomy of putative determinants of selective reporting in science based on saturated qualitative analyses of a representative sample of the relevant literature. The taxonomy clusters the determinants in a meaning-ful way. It consists of 12 mutually exclusive categories along with descriptions and examples to clarify boundaries and differences between the categories. The taxonomy should give structure and depth to commonly used expressions such as significance chasing (placed in the category *focus on preferred findings*) and conflict of interest (financial: *dependence upon sponsors*; intellectual: *relation-ship or collaboration issues* and *academic publication system hurdles*).

Two categories, *focus on preferred findings* and *poor or flexible research design*, covered over half of the determinants we found. These related mostly to choices of individual researchers or teams. The individual or team level was also referred to in six of the 10 other categories (*prejudice, dependence upon sponsors, lack of resources including time, doubts about reporting being worth the effort, limita-tions in reporting and editorial practices*-which refers to individual editors and authors*,* and *relationship and collaboration issues).* Four categories referred to the wider environment *(academic publication system hurdles, potential harm,* and two that included determinants with often no clear direction or hypothesis (e.g. when a range of disciplines or countries are compared based on distribu-tions): *high-risk area and its development*, and *unfavorable geographic or regula-tory environment.*

A recently developed framework of non-publication of clinical trials[19] attempted to answer the questions "what?" (defining the type of dissemination) "who?"(is to blame-actor/stakeholder) and "why?" (stakeholders' motivations). The basis of their framework comprised 50 highly cited articles published until 2012 identi-fied in Web of Science and consensus among 10 experts. Our work represents a wider scope of literature. For example, potential harm through bioterrorism was identified through veterinary medicine literature[36]. In addition, Web of Science

identified less than half of the articles in our sample. We included expert views and aggregate understanding, and determinants that may not have been studied well, yet in a different manner: through analyzing editorials, comments, and the full articles including introduction and discussion sections. We used explicit and transparent inductive qualitative research methods to cover the broad range of putative determinants in the literature. In contrast, Bassler et al.[19] focused on actors and motivations, which complements our work to help understand the multi-causality and multiple system pressures on and rewards for individuals and teams.

The combined quantitative and qualitative approaches including two different ways–prospectively and retrospectively–to check saturation increased the likelihood of having captured all relevant categories of determinants and served as an internal validation of our approach. However, the data did not suffice to discern patterns of determinants by academic field or strength of evidence. In our work, discussion to reach consensus was essential, as shown by modest initial agreement (72% and 78%) about inclusion of articles in reviewing abstracts and full-text.

Not publishing research outcomes is unethical. Our findings, however, raise questions about possible rare but legitimate reasons to report selectively or to not publish research. Obviously, potential harm should be considered a legitimate reason, when publication involves misuse by e.g. terrorists, or involves breaking confidentiality restrictions. Fatally flawed research probably should also not be published. However, poor design is preferably prevented in the first place, the academic reviewing system is in place to improve quality and to prevent fatally flawed work to be published or to be included in reviews and meta-analyses[37]. Future guidance may clarify what should be published in such cases.

Considering how categories were related (closer or farther apart, or relationships in the six cases of interaction we identified) helped conceptualize possible causal pathways. We used such findings to relate the 12 categories of determinants to propose a theory about possible causal pathways that lead to selective publication or reporting bias, following the principles of synthesis of qualitative data[31]. We hypothesize a causal mechanism in figure 3.3 which relates and orders the 12 determinant categories. We present four groups (A-D) through two higher-level classifications. The first classification is based on interpretation of how much influence or power individuals may exert in terms of their motivations and the means at their disposal (left side/dark blue versus right side/light blue of figure 3.3). The second classification is based on theoretical consideration of type of cause: necessary and component causes (bottom and top of figure 3.3)[38].
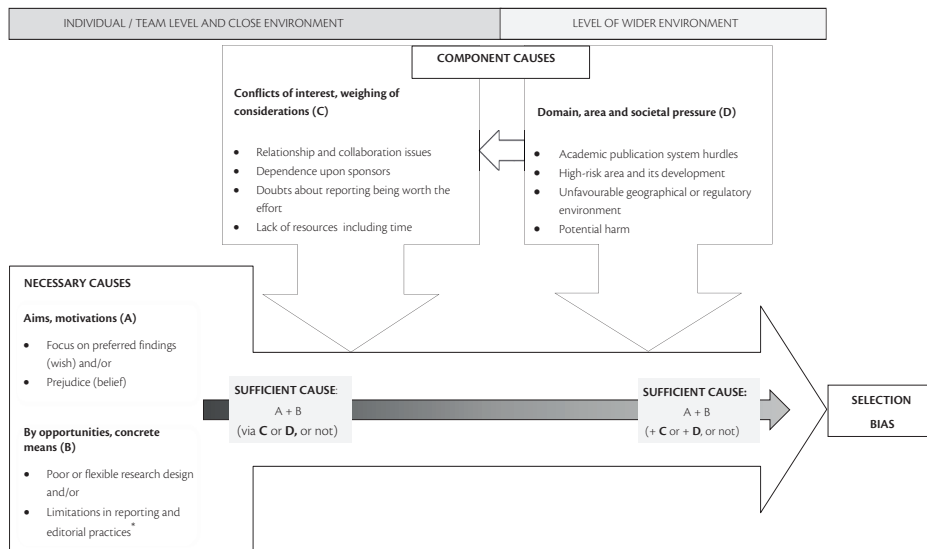
FIGURE 3.3 The determinant categories and possible causal pathways to selective reporting that results in selection bias. Bullet points indicate the 12 categories of determinants of selective reporting shown in Table 3.3 subsumed under four higher-level categories, A, B, C, D. *Refers mainly to reporting within single publications and to non-publication due to editorial decisions. Note that this graph implies interaction; effect modification by A and B (necessary causes) because there will be no selection bias with A or B alone. Interaction (or mediation) may also occur through C or D affecting A and B. However, C and D may shape aims/motivations (A) and opportunities/means (B) rather than fully controlling individuals' decisions.

The combined two most common categories, *focus on preferred findings*, and *poor or flexible research design,* suffice to cause bias through selective reporting. This is consistent with Ioannidis' statement that flexibility in designs, definitions, outcomes, and analytical modes increases the potential for transforming what would be "negative" results into "positive" results[16]. A sufficient cause represents an individual who is motivated and has the opportunity (means) to report selectively. Wishes and beliefs–the personal aims that motivate the individual–are quite different concepts although they come together in "wishful thinking"[39] also observed in motivated reasoning around interpretation of scientific findings to serve political interests[40,41]. The wishes or beliefs may or may not relate to intentional behavior and persons may or may not be fully aware of it[20]. Obviously, researchers and editors are key stakeholders because commonly they have decisional authority of what is actually being reported or published. Although each operates in the context of relationships and systems, it can be argued that researchers are the most important party because a single editor's decision is not decisive for non-publication in the scientific domain. In case of a series of rejections and submissions, it comes to researchers' *doubts about reporting being worth the effort* or *lack of resources including time* as component causes. At the

root may be a basic human attitude, the "very natural tendency to publish our successes"[42], with success, regardless how defined, driving direction of selective reporting. The pertinence of the second necessary factor–multiple opportunities to choose selectively what to analyze or report–is illustrated by at least 34 degrees of freedom that researchers have (in performing psychological research, most of which will apply to e.g., biomedical research as well)[43], and by accounts of selective reporting of positive findings even among trials registered at clinicaltrials.gov[44].

The necessary causes thus represent a motive (focus, prejudice; A in figure 3.3) and a means (design; B in figure 3.3), but there are other, component causes (C and D in figure 3.3). For example, important news is selectively reported but what is important news is shaped by the development within a scientific domain[35,45]. In addition, researchers' collaborations or relations with sponsors may nudge them to selectively report the preferred wishes of others. Beforehand, we expected a central role for a focus on preferred findings. However, a means or opportunity is necessary as well. After hypothesizing causal pathways in figure 3.3, we identified a relevant parallel with the Desire-Belief-Opportunity model that explains phenomena in sociology from Hedström[46]. This indicates that reconciling insights from various scientific domains such as psychology, sociology, economy and epidemiology is helpful.

New research, using various methods, should verify the categories we created and their interrelations as visualized with causal pathways that form a preliminary theory. The strengths of the proposed associations might be determined in a series of future studies. Future work may also refine the theory to increase relevance for specific disciplinary fields (e.g., non-clinical biomedical research). Nevertheless, because the causal pathways hold face validity and are consistent with theories developed in the social domain, we feel that our work can already help to design further research on the effectiveness of interventions. Such (complex) interventions should address the determinant categories we identified. So far, most empirical work has been performed on poor or flexible research design, but not all findings refer to clear causes and therefore inform interventions (such as studies examining association of sample size with selective reporting of positive findings). Future research should also employ qualitative methods to address researchers' daily decision making and balancing of interests to better understand causal mechanisms and the multiple factors involved.

The taxonomy and the theory may also help plan studies on risk profiling (e.g. research domains in which flexible designs are commonly used, or where a particular mission prevails) which in turn may inform efficient policy development on responsible conduct of research. We hope our work will promote a constructive

debate on causes of reporting bias and contribute to decrease the mostly deleterious phenomenon of selective reporting in modern science.

SUPPLEMENTARY RESULTS

TABLE S3.1 References to the 64 articles included in the determinant analysis, per category

Determinant classification, category	References to source articles (see list below)
1. Focus on preferred findings	1-52
2. Poor or flexible research design	1,4,6,7,9,10,12-19,21-25,27-32,38,40,41,44,48,49,53-55
3. High-risk area and its development	3-5,7,9-12,19,21,23,25,27,29,30,32,41,45,47,49,53-59
4. Dependence upon sponsors	1,3,4,13,16,19,21,22,24,25,27,31,33,36,38,40,49-51,60
5. Prejudice	3,8,19,22-24,27,29,31,32,40,41,45,49,53,61,62
6. Lack of resources including time	1,19,21,23,36,45,50,53,55,59,63
7. Doubts about reporting being worth the effort	1,16,30,45,47,49,50,59
8. Limitations in reporting and editorial practices	2,7,10,21,23,33
9. Academic publication system hurdles	7,9,23,25,36,47,49,53
10. Unfavorable geographical or regulatory environment[§]	1,3,9,10,19,25,31,38,60
11. Relationship and collaboration issues	4,24,34,38,45,63
12. Potential harm	6,64
(13) Not specified	13,19,22,34,45,63

LIST OF REFERENCES TO THE ANALYZED LITERATURE

1.   Chapman SJ, Shelton B, Mahmood H, Fitzgerald JE, Harrison EM, Bhangu A. Discontinuation and non-publication of surgical randomised controlled trials: observational study. *BMJ.* 2014;349:g6870.

2.   Dal-Re R, Caplan AL. Time to ensure that clinical trial appropriate results are actually published. *Eur J Clin Pharmacol.* Apr 2014;70(4):491-493.

3.   Dieterich S, Keall PJ, Orton CG. Point/Counterpoint. Increasing dependence on industry-funded research creates higher risk of biased reporting in medical physics. *Med Phys.* Oct 2013;40(10):100601.

4.   Ferguson CJ, Brannick MT. Publication bias in psychological science: prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychol Methods.* Mar 2012;17(1):120-128.

5.   Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA.* Sep 2 2009;302(9):977-984.

6.   Mattila T, Stoyanova V, Elferink A, Gispen-de Wied C, de Boer A, Wohlfarth T. Insomnia medication: do published studies reflect the complete picture of efficacy and safety? *Eur Neuropsychopharmacol.* Jul 2011;21(7):500-507.

7.   Onishi A, Furukawa TA. Publication bias is underreported in systematic reviews published in high-impact-factor journals: metaepidemiologic study. *J Clin Epidemiol.* Dec 2014;67(12):1320-1326.

8.   Park IU, Peacey MW, Munafo MR. Modelling the effects of subjective and objective decision making in scientific peer review. *Nature.* Feb 06 2014;506(7486):93-96.

9.   Salami K, Alkayed K. Publication bias in pediatric hematology and oncology: analysis of abstracts presented at the annual meeting of the American Society of Pediatric Hematology and Oncology. *Pediatr Hematol Oncol.* Apr 2013;30(3):165-169.

10.   Sando IC, Malay S, Chung KC. Analysis of publication bias in the literature for distal radius fracture. *J Hand Surg Am.* May 2013;38(5):927-934 e925.

11.   Simundic AM. Bias in research. *Biochem Med (Zagreb).* 2013;23(1):12-15.

12.   Smulders YM. A two-step manuscript submission process can reduce publication bias. *J Clin Epidemiol.* Sep 2013;66(9):946-947.

13.   Sune P, Sune JM, Montoro JB. Positive outcomes influence the rate and time to publication, but not the impact factor of publications of clinical trial results. *PLoS One.* 2013;8(1):e54583.

14.   Trinquart L, Ioannidis JP, Chatellier G, Ravaud P. A test for reporting bias in trial networks: simulation and case studies. *BMC Med Res Methodol.* Sep 27 2014;14:112.

15.   Tzoulaki I, Siontis KC, Evangelou E, Ioannidis JP. Bias in associations of emerging biomarkers with cardiovascular disease. *JAMA Intern Med.* Apr 22 2013;173(8):664-671.

16.   Abaid LN, Grimes DA, Schulz KF. Reducing publication bias through trial registration. *Obstet Gynecol.* Jun 2007;109(6):1434-1437.

17.   Anderson HR, Atkinson RW, Peacock JL, Sweeting MJ, Marston L. Ambient particulate matter and health effects: publication bias in studies of short-term associations. *Epidemiology.* Mar 2005;16(2):155-163.

18.   Bardy AH. Bias in reporting clinical trials. *Br J Clin Pharmacol.* Aug 1998;46(2):147-150.

19. Begg CB, Berlin JA. Publication bias and dissemination of clinical research. *J Natl Cancer Inst.* Jan 18 1989;81(2):107-115.

20. Chalmers I. From optimism to disillusion about commitment to transparency in the medico-industrial complex. *J R Soc Med.* Jul 2006;99(7):337-341.

21. Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ.* Apr 2 2005;330(7494):753.

22. Dickersin K, Min YI. NIH clinical trials and publication bias. *Online J Curr Clin Trials.* Apr 28 1993;Doc No 50:[4967 words; 4953 paragraphs].

23. Gilbert JR, Williams ES, Lundberg GD. Is there gender bias in JAMA's peer review process? *JAMA.* Jul 13 1994;272(2):139-142.

24. Greenland S. Accounting for uncertainty about investigator bias: disclosure is informative. *J Epidemiol Community Health.* Aug 2009;63(8):593-598.

25. Harris IA, Mourad M, Kadir A, Solomon MJ, Young JM. Publication bias in abstracts presented to the annual meeting of the American Academy of Orthopaedic Surgeons. *J Orthop Surg (Hong Kong).* Apr 2007;15(1):62-66.

26. Hopewell S, Clarke M, Stewart L, Tierney J. Time to publication for results of clinical trials. *Cochrane Database Syst Rev.* Apr 18 2007(2):MR000011.

27. Ioannidis JP. Why most published research findings are false. *PLoS Med.* Aug 2005;2(8):e124.

28. Nassir Ghaemi S, Shirzadi AA, Filkowski M. Publication bias and the pharmaceutical industry: the case of lamotrigine in bipolar disorder. *Medscape J Med.* 2008;10(9):211.

29. Newcombe RG. Towards a reduction in publication bias. *Br Med J (Clin Res Ed).* Sep 12 1987;295(6599):656-659.

30. Okike K, Kocher MS, Mehlman CT, Heckman JD, Bhandari M. Publication bias in orthopaedic research: an analysis of scientific factors associated with publication in the Journal of Bone and Joint Surgery (American Volume). *J Bone Joint Surg Am.* Mar 2008;90(3):595-601.

31. Olson CM, Rennie D, Cook D, et al. Publication bias in editorial decision making. *JAMA.* Jun 05 2002;287(21):2825-2828.

32. Preston C, Ashby D, Smyth R. Adjusting for publication bias: modelling the selection process. *J Eval Clin Pract.* May 2004;10(2):313-322.

33. Rampton S, Stauber J. Research funding, conflicts of interest, and the meta-methodology of public relations. *Public Health Rep.* Jul-Aug 2002;117(4):331-339.

34. Rifai N, Altman DG, Bossuyt PM. Reporting bias in diagnostic and prognostic studies: time for action. *Clin Chem.* Jul 2008;54(7):1101-1103.

35. Steinbrook R. Public registration of clinical trials. *N Engl J Med.* Jul 22 2004;351(4):315-317.

36. Stewart LA, Parmar MK. Bias in the analysis and reporting of randomized controlled trials. *Int J Technol Assess Health Care.* Spring 1996;12(2):264-275.

37. Takeda A, Loveman E, Harris P, Hartwell D, Welch K. Time to full publication of studies of anti-cancer medicines for breast cancer and the potential for publication bias: a short systematic review. *Health Technol Assess.* Oct 2008;12(32):iii, ix-x, 1-46.

38. von Elm E, Rollin A, Blumle A, Huwiler K, Witschi M, Egger M. Publication and non-publication of clinical trials: longitudinal study of applications submitted to a research ethics committee. *Swiss Med Wkly.* Apr 5 2008;138(13-14):197-203.

39. The time is coming. *Nat Rev Drug Discov.* Nov 2004;3(11):897-898.

40.   de Melker HE, Rosendaal FR, Vandenbroucke JP. [The importance of publication bias in medical-scientific literature]. *Ned Tijdschr Geneeskd.* Oct 16 1993;137(42):2126-2130.

41.   Ferguson CJ, Heene M. A Vast Graveyard of Undead Theories: Publication Bias and Psychological Science's Aversion to the Null. *Perspect Psychol Sci.* Nov 2012;7(6):555-561.

42.   Guemann M. Registration of clinical trials in physiotherapy, an ISPJE's recommendation to fight against the publication biases. *Kinesitherapie.* 2013;13(137):1.

43.   Hoye A, Elvik R. Publication Bias in Road Safety Evaluation How Can It Be Detected and How Common Is It? *Transp. Res. Record.* 2010(2147):1-8.

44.   Lang L. New Study Finds Publication Bias Among Trials Submitted to the US Food and Drug Administration. *Gastroenterology.* 2009;136(2):374.

45.   McCambridge J. A response to the commentaries: look away now or face up to the profound problem of publication bias in drug education research. *Drug Alcohol Rev.* 2008;27(4):352-356.

46.   McDaniel MA, Rothstein HR, Whetzel DL. Publication bias: A case study of four test vendors. *Pers. Psychol.* Win 2006;59(4):927-953.

47.   Parmley WW. Publication bias. *J Am Coll Cardiol.* Nov 01 1994;24(5):1424-1425.

48.   Rothstein HR, Bushman BJ. Publication bias in psychological science: comment on Ferguson and Brannick (2012). *Psychol Methods.* Mar 2012;17(1):129-136.

49.   Shields PG. Publication bias is a scientific problem with adverse ethical outcomes: the case for a section for null results. *Cancer Epidemiol Biomarkers Prev.* Aug 2000;9(8):771-772.

50.   Song F, Hooper L, Loke YK. Publication bias: What is it? How do we measure it? How do we avoid it? *Reports in Medical Imaging.* 2012;5(1):71-81.

51.   Bardy AH. Report bias in drug research. *Therapie.* Jul-Aug 1996;51(4):382-383.

52.   Reidenberg MM. Decreasing publication bias. *Clin Pharmacol Ther.* Jan 1998;63(1):1-3.

53.   Goldsmith LA, Blalock EN, Bobkova H, Hall RP, 3rd. Picking your peers. *J Invest Dermatol.* Jul 2006;126(7):1429-1430.

54.   Moher D, Fortin P, Jadad AR, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet.* Feb 10 1996;347(8998):363-366.

55.   Negrini S. Another form of publication bias: the unpublished "everyday" clinical researches. *Eur J Phys Rehabil Med.* Dec 2011;47(4):531-532.

56.   Altman DG. Poor-quality medical research: what can journals do? *JAMA.* Jun 05 2002;287(21):2765-2767.

57.   Hopewell S, McDonald S. Full publication of trials initially reported as abstracts in the Australian and New Zealand Journal of Medicine 1980-2000. *Intern Med J.* Apr 2003;33(4):192-194.

58.   Munafo MR, Flint J. How reliable are scientific studies? *Br J Psychiatry.* Oct 2010;197(4):257-258.

59.   Sherrod J. Selective publication of information. *Spec. Libr.* 1964;55(6):386-387.

60.   Bourgeois FT, Murthy S, Mandl KD. Outcome reporting among drug trials registered in ClinicalTrials.gov. *Ann Intern Med.* Aug 03 2010;153(3):158-166.

61.   Fisher M, Friedman SB, Strauss B. The effects of blinding on acceptance of research papers by peer review. *JAMA.* Jul 13 1994;272(2):143-146.

62.   Pless B. Are Editors free from bias? The special case of Letters to the Editor. *Inj Prev.* Dec 2006;12(6):353-354.

3

63. Dowd MD. Breaching the contract: the ethics of nonpublication of research studies. *Arch Pediatr Adolesc Med.* Oct 2004;158(10):1014-1015.

64. Gyles C. Is there ever good reason to not publish good science? *Can Vet J.* Jun 2012;53(6):587-588.

3

REFERENCES

1. Hartmann M. Impact assessment of the European Clinical Trials Directive: a longitudinal, prospective, observational study analyzing patterns and trends in clinical drug trial applications submitted since 2001 to regulatory agencies in six EU countries. *Trials.* Apr 29 2012;13:53.

2. Dickersin K, Chalmers I. Recognizing, investigating and dealing with incomplete and biased reporting of clinical research: from Francis Bacon to the WHO. *J R Soc Med.* Dec 2011;104(12):532-538.

3. Bakobaki JM, Rauchenberger M, Joffe N, McCormack S, Stenning S, Meredith S. The potential for central monitoring techniques to replace on-site monitoring: findings from an international multi-centre clinical trial. *Clin Trials.* Apr 2012;9(2):257-264.

4. Moorthy VS, Karam G, Vannice KS, Kieny MP. Rationale for WHO's new position calling for prompt reporting and public disclosure of interventional clinical trial results. *PLoS Med.* Apr 2015;12(4):e1001819.

5. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet.* Jul 04 2009;374(9683):86-89.

6. Jones CW, Handler L, Crowell KE, Keil LG, Weaver MA, Platts-Mills TF. Non-publication of large randomized clinical trials: cross sectional analysis. *BMJ.* 2013;347:f6104.

7. Ross JS, Mulvey GK, Hines EM, Nissen SE, Krumholz HM. Trial publication after registration in ClinicalTrials.Gov: a cross-sectional analysis. *PLoS Med.* Sep 2009;6(9):e1000144.

8. Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess.* Feb 2010;14(8):iii, ix-xi, 1-193.

9. Macleod MR, Lawson McLean A, Kyriakopoulou A, et al. Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. *PLoS Biol.* Oct 2015;13(10):e1002273.

10. Sena ES, van der Worp HB, Bath PM, Howells DW, Macleod MR. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol.* Mar 30 2010;8(3):e1000344.

11. Sterling TD. Publication decisions and Their Possible Effects on Inferences Drawn from Tests of Significance--Or Vice Versa. *Journal of the American Statistical Association.* 1959;54, No. 285 (Mar. 1959):30-34.

12. Decullier E, Lheritier V, Chapuis F. Fate of biomedical research protocols and publication bias in France: retrospective cohort study. *BMJ.* Jul 2 2005;331(7507):19.

13. Petticrew M, Egan M, Thomson H, Hamilton V, Kunkler R, Roberts H. Publication bias in qualitative research: what becomes of qualitative research presented at conferences? *J Epidemiol Community Health.* Jun 2008;62(6):552-554.

14. Cooper H, DeNeve K, Charlton K. Finding the missing science: The fate of studies submitted for review by a human subjects committee. *Psychol Methods.* Dec 1997;2(4):447-452.

15. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev.* Jan 21 2009(1):MR000006.

16. Ioannidis JP. Why most published research findings are false. *PLoS Med.* Aug 2005;2(8):e124.

17. Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. *Cochrane Database Syst Rev.* Feb 16 2017;2:MR000033.

18. McGauran N, Wieseler B, Kreis J, Schuler YB, Kolsch H, Kaiser T. Reporting bias in medical research - a narrative review. *Trials*. 2010;11:37.

19. Bassler D, Mueller KF, Briel M, et al. Bias in dissemination of clinical research findings: structured OPEN framework of what, who and why, based on literature review and expert consensus. *BMJ Open*. Jan 21 2016;6(1):e010024.

20. Greenland S. Accounting for uncertainty about investigator bias: disclosure is informative. *J Epidemiol Community Health*. Aug 2009;63(8):593-598.

21. Dunkelberg S. Further information on publication bias. *Zeitschrift fur Allgemeinmedizin*. 2005;81(4):145-146.

22. Martinson BC, Anderson MS, Crain AL, de Vries R. Scientists' perceptions of organizational justice and self-reported misbehaviors. *J Empir Res Hum Res Ethics*. Mar 2006;1(1):51-66.

23. Fanelli D. Do pressures to publish increase scientists' bias? An empirical support from US States Data. *PLoS One*. Apr 21 2010;5(4):e10271.

24. Davis C, Lexchin J, Jefferson T, Gotzsche P, McKee M. "Adaptive pathways" to drug authorisation: adapting to industry? *BMJ*. Aug 16 2016;354:i4437.

25. Greenhalgh T, Taylor R. Papers that go beyond numbers (qualitative research). *BMJ*. Sep 20 1997;315(7110):740-743.

26. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nurs Health Sci*. Sep 2013;15(3):398-405.

27. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. Jan 01 2015;4:1.

28. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*. Jan 02 2015;349:g7647.

29. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. Jul 21 2009;6(7):e1000097.

30. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med*. Sep 2014;89(9):1245-1251.

31. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Med Res Methodol*. Nov 27 2012;12:181.

32. Elo S, Kaariainen M, Kanste O, Polkki T, Utriainen K, Kyngas H. Qualitative Content Analysis: A Focus on Trustworthiness. *SAGE Open* 2014;4(1).

33. Fisher M, Friedman SB, Strauss B. The effects of blinding on acceptance of research papers by peer review. *JAMA*. Jul 13 1994;272(2):143-146.

34. Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ*. Apr 2 2005;330(7494):753.

35. Preston C, Ashby D, Smyth R. Adjusting for publication bias: modelling the selection process. *J Eval Clin Pract*. May 2004;10(2):313-322.

36. Gyles C. Is there ever good reason to not publish good science? *Can Vet J*. Jun 2012;53(6):587-588.

37. Ferguson CJ, Brannick MT. Publication bias in psychological science: prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychol Methods*. Mar 2012;17(1):120-128.

38. Rothman KJ. Causes. *Am J Epidemiol.* Dec 1976;104(6):587-592.

39. Bastardi A, Uhlmann EL, Ross L. Wishful thinking: belief, desire, and the motivated evaluation of scientific evidence. *Psychol Sci.* Jun 2011;22(6):731-732.

40. Colombo M, Bucher L, Inbar Y. Explanatory Judgment, Moral Offense and Value-Free Science. *Rev Philos Psychol.* 2016;7(4):743-763.

41. Kraft PW, Lodge M, Taber CS. Why People "Don't Trust the Evidence": Motivated Reasoning and Scientific Beliefs. *Ann. Am. Acad. Polit. Soc. Sci.* Mar 2015;658(1):121-133.

42. Dickersin K, Min YI. NIH clinical trials and publication bias. *Online J Curr Clin Trials.* Apr 28 1993.

43. Wicherts JM, Veldkamp CL, Augusteijn HE, Bakker M, van Aert RC, van Assen MA. Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. *Front Psychol.* 2016;7:1832.

44. Dechartres A, Bond EG, Scheer J, Riveros C, Atal I, Ravaud P. Reporting of statistically significant results at ClinicalTrials.gov for completed superiority randomized controlled trials. *BMC Med.* Nov 30 2016;14(1):192.

45. Ter Riet G, Bouter LM. *Animal Models for Human Cancer: Discovery and Development of Novel Therapeutics, Volume 69. Chapter 5: How to End Selective Reporting in Animal Research (p. 61-77)*2016.

46. Hedström P. *Dissecting the Social - On the Principles of Analytical Sociology.* Cambridge Cambridge University Press; 2005.

3

Chapter 4

Clinical drug trials in the context of globalized drug development

Chapter 4.1

Contrasting clinical trials in oncology and psychiatry: landscape as proxy for efficiency in drug development

Cornelis A van den Bogert, Hubert GM Leufkens

ABSTRACT

In the field of oncology, many new drugs have reached the market over the past years. In contrast, the field of psychiatry has been increasingly abandoned by drug developers. Phase 3 clinical trials are the key in bringing new drugs to the market, and in improving existing therapies. To gain insight in the actual late-phase clinical drug development activity in these fields over the past years in the EU, we contrasted characteristics of phase 3 clinical drug trials between oncology and psychiatry. We used the EudraCT database to select all phase 3 clinical drug trials reviewed by the national competent authorities within the EU between 2013 and 2015 (n = 2255). Of these, we further analyzed the oncology (n = 471) and psychiatry (n = 65) phase 3 trials. A similar proportion of the oncology and psychiatry trials was industry-sponsored (63.2% vs. 58.5%, respectively). Oncology trials were relatively more often multicenter (93.8%) compared to psychiatry trials (75.4%). Besides the lack of understanding the biology and etiology of psychiatric disorders, the low phase 3 trial activity in psychiatry may be explained by the controversy surrounding diagnosis and medicalization of psychiatric disorders. Another difference between drug development in oncology and psychiatry is the availability of biomarkers that can predict efficacy and safety. If available, implementation of these biomarkers in the design of phase 3 trials could be stimulated, to implement targeted therapy as early as possible. Finally, quality of life is an important, but yet still underused endpoint of clinical trials in both disease areas. The new Clinical Trial Regulation should aim to facilitate these aspects to sustain the EU as vital area for clinical drug development.

INTRODUCTION

There have been many analyses and reflections on why drug development for cancer have been so successful over the last decade [1-5], and why psychiatry is almost at the other side of the spectrum with virtually no advances in bringing new innovative products to the clinic [6-10].

Key to the success in oncology was the identification of several pathophysiological drivers of the disease, and the translation of these drivers to drug targets that correlated well with prolonged clinical survival[11]. These principles of targeting specific disease drivers and objectively measureable endpoints has delivered a proven track record of authorized drugs against a wide range of tumors [1]. The major successes in terms of clinical advances and revenue generation of oncology products contributed to the dominance of oncology in the current clinical trial landscape[2]. Clinical cancer research institutes and the pharmaceutical industry are working intensively together to fill the knowledge gaps based on their clinical experience and investigate new combination treatment and dosage regimens.

In contrast, the psychiatrist has seen only but few new drug therapies been added to his therapeutic arsenal over the past decades [12]. Unlike in oncology, the etiology of psychiatric disorders has remained largely unclear [12]. The lack of druggable mechanistic drivers of the disease that correlate well with endpoints partially explains the unsatisfactorily low output of the psychiatry development pipeline[11]. This inherently relates to the complex, multifaceted nature of psychiatric disorders in which "all or none" effects cannot be expected, such as with tumor drugs targeted at one receptor. In addition, there are currently no biomarkers validated to predict an effect on clinical outcomes [13]. Moreover, the concern on geographical variability and reliability of both trial outcomes and diagnostics to select the right population are also barriers to efficient clinical development, which is nowadays conducted in a global context[14-16].

The sustained success in oncology is not just the result of ongoing breakthrough scientific discoveries. In fact, many of the new cancer drugs are, on the molecular level, incremental improvements of existing drugs. For example, the difference between the tyrosine kinase inhibitors sorafenib (Nexavar) and regorafenib (Stivarga) is the replacement of a hydrogen atom for a fluorine atom [17]. This type of innovation of cancer therapies is generally seen as positive, as the oncologist gets more options to treat complex and therapy-resistant tumors. In psychiatry, this type of innovation has also occurred. Several variants of the atypical antipsychotics and selective serotonin reuptake inhibitors have been developed. Contrary to oncology however, these drugs were often received with a critical attitude and labeled as "me-too" drugs[18]. More broadly stated, drug development in psychiatry

has always been surrounded by debate and controversy as to whether psychiatric problems such as schizophrenia should be viewed and diagnosed as disease, and should be treated with drugs[19-22].

Thus, in sharp contrast with oncology, there is no consensus in the field of psychiatry on the position of pharmacotherapy in treating patients with mental problems, which has made the field unattractive for commercial drug development. However, also post-marketing clinical trial activity (trials on authorized products) is relatively lower in psychiatry (69% of the psychiatry trials were on authorized products) compared to oncology (74% of the oncology trials were on authorized products). These trials do not require improved understanding of the disease biology, or the prospect of bringing a new product to the market, but rather public and private incentives to conduct them. Repurposing existing drugs (also from other disease areas) in psychiatry can be quite successful. For example, the old antihypertensive drug clonidine was authorized in 2010 for the treatment of attention-deficit hypersensitivity disorder [23]. Furthermore, the absence of biology-based biomarkers [13,24] does not mean that targeted therapy cannot be a part of psychiatry trials. In phase 3 trials, patients can be stratified in subgroups based on their symptoms, across different indications, which may target drugs toward their optimal target population in preferably a head-to-head design [25-27].

For sure, addressing the question why one clinical field is so successful and another not, will not deliver a clear-cut answer. Particularly because the history of drug development shows that there are time trends as well. Certain therapeutic fields have benefited from new science and a better understanding of disease and underlying etiology, and hence to identification of druggable targets and new drugs. This is illustrated by the discovery of the aspirin as a suppressor of prostaglandin and thromboxane, which directly led to the clinical application of aspirin as platelet inhibitor against thrombosis [28,29]. Other examples of successful drug development and pharmacotherapeutic innovation following scientific discoveries include the exploitation of the renin angiotensin aldosterone system[30] and drug treatment against retroviral infections [31].

Whatever explanation one may propose for differences in progress in therapeutic advances across clinical domains, clinical trials, particularly randomized controlled phase 3 trials, remain the powerhouse in bringing medicines with proven efficacy and an acceptable safety profile to the clinic [32]. Phase 3 trials can test novel drugs from the pipeline of companies, but also existing drugs against new diseases, in new combinations, or new dosages.

By contrasting the characteristics of clinical trial activity in oncology with psychiatry, two very different fields in terms of diseases and scientific discover-

ies, valuable learning points may be extracted to enhance clinical trial activity in both fields. Furthermore, the EU area is currently facing the new clinical trial regulation, of which the implementation is scheduled for the third quarter of 2018[33]. To enable evaluation of the impact of the new regulation, a baseline measurement of the situation before its implementation is helpful. Therefore, in this analysis, we aimed to provide insights in clinical phase 3 drug research in oncology and psychiatry by describing and contrasting key characteristics of the phase 3 clinical drug trial landscape in the EU.

METHODS

We looked at all phase 3 drug trials in the areas of oncology (testing drugs against any tumor) and psychiatry (testing drugs against any mood or behavioral disorder) reviewed by national competent authorities across the EU and registered in the EudraCT database between 2013 and 2015. EudraCT is the database established by the European Union[34], maintained by the European Medicines Agency (EMA), in which all clinical drug trial applications (CTAs) are registered and in which the national competent trial authorities (NCAs) keep track of the CTA review process. The phase 3-trials in EudraCT are publicly accessible through the EU Clinical Trial Register[34]. On our request, the EMA approved the extraction of a data report containing all phase 3 trials reviewed by the NCAs in the EU registered in EudraCT between 2013 and 2015. The NCA of the Netherlands (the Central Committee on Research Involving Human Subjects, CCMO) facilitated the extraction. The report contained information about the medical condition under investigation (recoded as the disease area according to the International Classification of Diseases (ICD) version 10); the sponsor (recoded as industry-sponsored versus investigator initiated); whether the trial was single-center, multicenter with sites only within the EU, or multicenter with sites both within and outside the EU; the estimated number of participants within the EU and for the whole trial; and whether the drug under investigation was authorized for marketing in the EU or in the countries where the trial was conducted.

RESULTS

Between 2013 and 2015, 2255 NCA-reviewed phase 3 clinical drug trials were registered in the EudraCT database with a decision of the national competent authority. Of these, 471 (20.1%) were in oncology and 65 (2.9%) were in psychia-

try, numbers shown in figure 4.1.1. The sum target number of participants to be included in the oncology trials was 257,765, of which 144,822 (56%) participants from within the EU. For psychiatry trials, the sum target sample size was 22,672, of which 14,016 (61.8%) from within the EU.
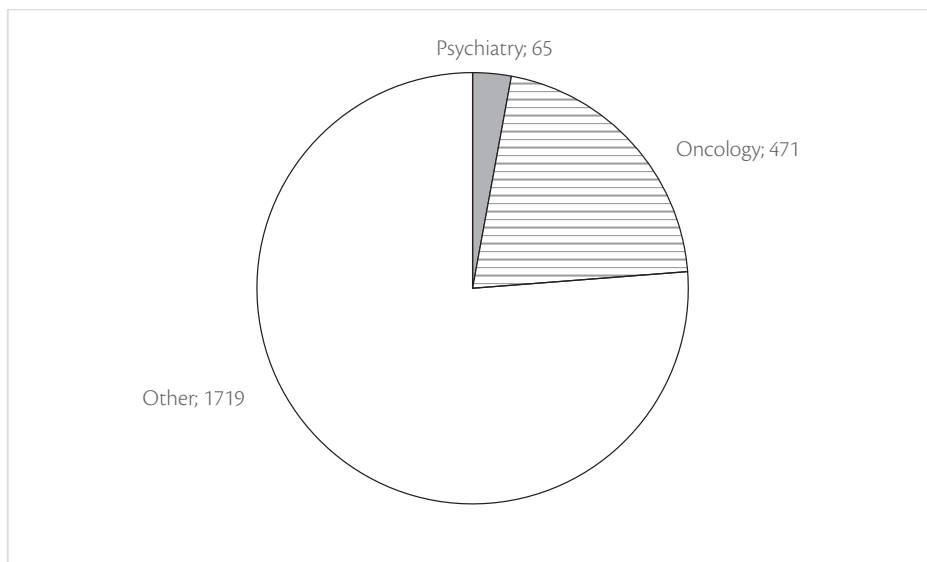
FIGURE 4.1.1 Phase 3 clinical drug trial application reviews by the national competent authorities within the European Union between 2013 and 2015.

The distribution of industry-sponsored and investigator-initiated trials was similar for psychiatry and oncology. Taken the 3 years together, 298/471 (63.2%) oncology trials were industry-sponsored and 173/471 (36.7%) were investigator-initiated, vs. 38/65 (58.5%) industry-sponsored trials and 27/65 (41.5%) investigator-initiated trials in psychiatry. Figure 4.1.2 shows that there was a drop in investigator-initiated oncology trials in 2015, but it is unknown whether this observation is an outlier or trend. As illustrated in figure 4.1.3, the higher clinical trial activity in oncology is mostly multicenter trials. Over the 3 years, 442/471 (93.8%) oncology trials were multicenter versus 49/65 (75.4%) multicenter trials in psychiatry. Figure 4.1.4 shows that 348/468 (74.4%) trials in oncology involved already authorized drug products versus 44/64 (68.9%) trials involving authorized drug products in psychiatry (in three oncology trials and in one psychiatry trial the authorization status of the products was not indicated).
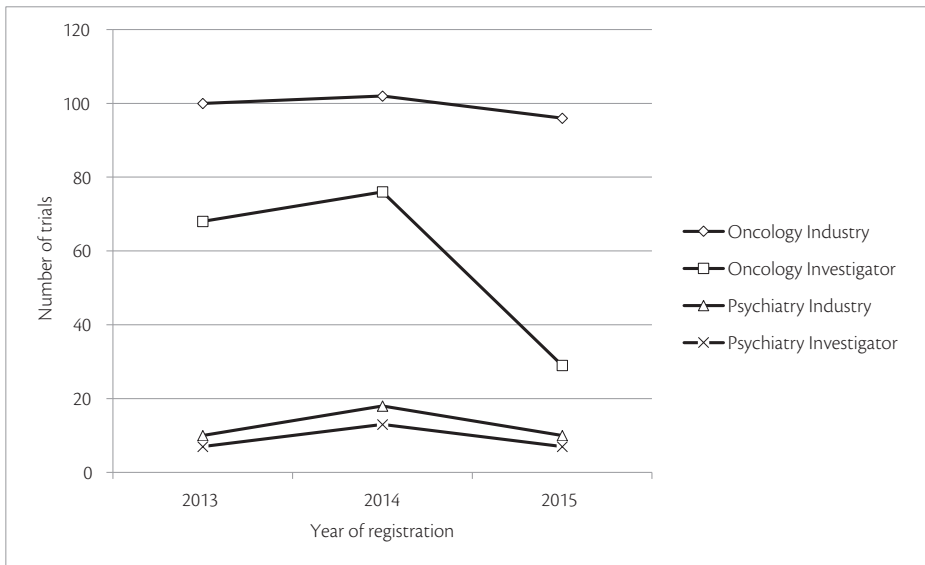
FIGURE 4.1.2 Number of oncology and psychiatry phase 3 drug trials, stratified by sponsor type.
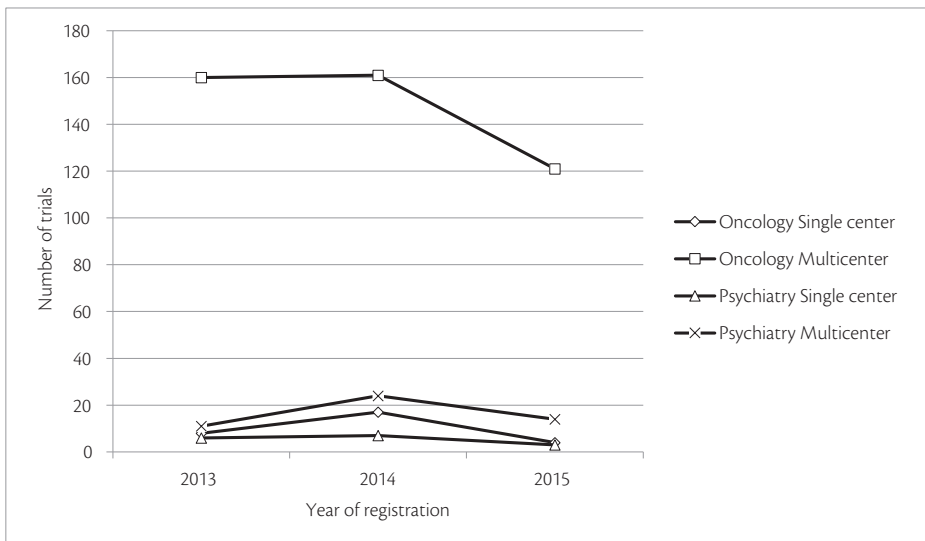
FIGURE 4.1.3 Number of oncology and psychiatry phase 3 drug trials, stratified by center(s).
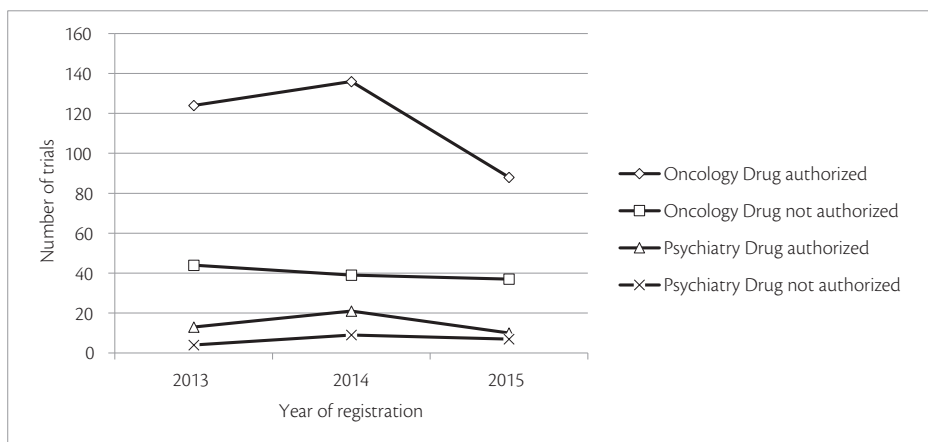
FIGURE 4.1.4 Number of oncology and psychiatry phase 3 drug trials, stratified by authorization status of trial drug.

DISCUSSION

Our data show that over the period of 2013-2015 the phase 3 trial activity in the EU seems to be in decreasing trend. This decrease may reflect a number of developments, including concentration and/or decrease of clinical R&D investments by the pharmaceutical industry, outsourcing of phase 3 clinical trials to countries outside the EU, and a lower availability of funding for larger investigator-initiated trials. To which extent the decrease is caused by regulatory requirements is unknown. Whether the new clinical trial regulation will have an impact, needs to be evaluated within a few years after implementation of the regulation.

The contrast between oncology and psychiatry is large in terms of absolute numbers of phase 3 trials, patients and hence investments in clinical R&D. Our analysis shows that the higher phase 3 trial activity in oncology in particular concerns multicenter trials, and trials on products that are already authorized. This can be explained by the major pharmacotherapeutic successes of oncology over the past decades, which has created a resource-rich clinical R&D infrastructure with both industry and government investments. Not only industry, but also investigators collaborate on the international level (through the European Organisation for Research and Treatment of Cancer, EORTC), to improve drugs already available in the clinic. In contrast, the absence of success in psychiatry has discouraged companies and investigators to invest in research on new psychiatric treatments. This decline in clinical research is viewed as undesirable by clinicians and policy makers [35-37], as the current unmet medical need of patients and society will unlikely resolve in this way. As clinical trial regulations have

been recognized as potential barriers to trials in psychiatry, it should be monitored whether the new regulation works out properly[37].

A balanced origin of trial populations, as well as clinical trials designed and conducted independently from the pharmaceutical industry, are both elements that the regulatory system should facilitate[38,39]. Sustaining this activity remains needed, through both industry R&D and grassroots academic initiatives. With the introduction of the new European regulation, which aims to boost responsible, sustainable and transparent conduct of clinical drug trials within the EU [40], it is important to evaluate if the changes in regulatory review have the desired outcome. This means that the initiation of academic drug trials should not be discouraged by more bureaucracy and regulations [41,42].

The introduction of biomarkers for targeted therapy is not without hurdles, as illustrated by the field of oncology. For example, the epidermal growth factor receptor inhibitor cetuximab, initially authorized in 2004, was restricted to KRAS-wildtype colon carcinoma by the EMA by the end of 2008. The incidence and relevance of KRAS mutations in (the pathophysiology of) colorectal tumors had already been demonstrated in the late 1990s[43,44]. If the KRAS biomarker had been incorporated in premarketing trials from the beginning, the ineffective treatment of 30% of the KRAS-mutated colon carcinomas could have been avoided. Therefore, even in the case of existing validated biomarkers, it is not guaranteed that they are used to select the optimal patient population. Clinical trial regulators should promote the inclusion of biomarkers in phase 3 clinical trials, so that their relevance is clear for doctors, patients and payers as early in the lifecycle as possible.

Another issue of concern in both fields is the priority of the evaluation of quality of life in clinical trials. Besides the measurements of survival (cancer) and mental status (psychiatry), the "*do I live better?*" question requires attention. In particular, when a treatment is tested against metastasized and incurable cancer, or persistent psychiatric illness, the question of the quality of life becomes important. If a patient is offered a drug therapy which may extend their life for several months, it should be known whether the treatment will also improve (or not worsen) their life, compared to supportive care. Measuring quality of life goes beyond standard toxicity measures or mood scales[45], and is a separate aspect that needs – and can – be reliably investigated [46-48]. As measuring quality of life requires usually a long follow-up, harmonizing ethical review timelines should streamline the conduct of follow-up studies on the quality of life of trial participants [49].

To conclude, we analyzed the phase 3 clinical drug trial activity within the EU in the fields of oncology and psychiatry, showing a large contrast between these

disease areas. Explanations of this contrast include the lack of pharmacothera-peutic successes and scientific progress, disagreement in medicine and society about the nature and treatment of psychiatric diseases, and lack of biology-based biomarkers. More post-marketing trials such as drug repurposing and head-to-head comparisons are needed in psychiatry. To stimulate and optimize drug development in psychiatry, the new clinical trial regulation in the EU should encourage clinical trials to incorporate targeted therapy and quality of life parameters in the design of trials in cancer and psychiatry.

ACKNOWLEDGEMENTS

REFERENCES

1   Aggarwal, S. Targeted cancer therapies. *Nat Rev Drug Discov* 9, 427-428, doi:10.1038/nrd3186 (2010).

2   Berggren, R., Moller, M., Moss, R., Poda, P. & Smietana, K. Outlook for the next 5 years in drug innovation. *Nat Rev Drug Discov* 11, 435-436, doi:10.1038/nrd3744 (2012).

3   Booth, B., Glassman, R. & Ma, P. Oncology's trials. *Nat Rev Drug Discov* 2, 609-610, doi:10.1038/nrd1158 (2003).

4   DiMasi, J. A. & Grabowski, H. G. Economics of new oncology drug development. *J Clin Oncol* 25, 209-216, doi:10.1200/JCO.2006.09.0803 (2007).

5   Reichert, J. M. & Wenger, J. B. Development trends for new cancer therapeutics and vaccines. *Drug Discov Today* 13, 30-37, doi:10.1016/j.drudis.2007.09.003 (2008).

6   Barbui, C. & Bighelli, I. A new approach to psychiatric drug approval in Europe. *PLoS Med* 10, e1001530, doi:10.1371/journal.pmed.1001530 (2013).

7   Carpenter, W. T. & Koenig, J. I. The evolution of drug development in schizophrenia: past issues and future opportunities. *Neuropsychopharmacology* 33, 2061-2079, doi:10.1038/sj.npp.1301639 (2008).

8   Conn, P. J. & Roth, B. L. Opportunities and challenges of psychiatric drug discovery: roles for scientists in academic, industry, and government settings. *Neuropsychopharmacology* 33, 2048-2060, doi:10.1038/sj.npp.1301638 (2008).

9   Pangalos, M. N., Schechter, L. E. & Hurko, O. Drug development for CNS disorders: strategies for balancing risk and reducing attrition. *Nat Rev Drug Discov* 6, 521-532, doi:10.1038/nrd2094 (2007).

10   Young, J. W. & Geyer, M. A. Developing treatments for cognitive deficits in schizophrenia: the challenge of translation. *J Psychopharmacol* 29, 178-196, doi:10.1177/0269881114555252 (2015).

11   Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat Rev Drug Discov* 16, 19-34, doi:10.1038/nrd.2016.230 (2017).

12   Chancellor, D. The depression market. *Nat Rev Drug Discov* 10, 809-810, doi:10.1038/nrd3585 (2011).

13   Scarr, E. *et al.* Biomarkers for Psychiatry: The Journey from Fantasy to Fact, a Report of the 2013 CINP Think Tank. *Int J Neuropsychopharmacol* 18, pyv042, doi:10.1093/ijnp/pyv042 (2015).

14   Welten, C. C. *et al.* Efficacy of drug treatment for acute mania differs across geographic regions: An individual patient data meta-analysis of placebo-controlled studies. *J Psychopharmacol* 29, 923-932, doi:10.1177/0269881115586938 (2015).

15   Chmielewski, M., Clark, L. A., Bagby, R. M. & Watson, D. Method matters: Understanding diagnostic reliability in DSM-IV and DSM-5. *J Abnorm Psychol* 124, 764-769, doi:10.1037/abn0000069 (2015).

16   Kraemer, H. C., Kupfer, D. J., Clarke, D. E., Narrow, W. E. & Regier, D. A. DSM-5: how reliable is reliable enough? *Am J Psychiatry* 169, 13-15, doi:10.1176/appi.ajp.2011.11010050 (2012).

17   Sviripa, V. *et al.* Fluorinated N,N'-diarylureas as AMPK activators. *Bioorg Med Chem Lett* 23, 1600-1603, doi:10.1016/j.bmcl.2013.01.096 (2013).

18   Becker, R. E., Seeman, M. V., Greig, N. H. & Lahiri, D. K. What can triumphs and tribulations from drug research in Alzheimer's disease tell us about the development

of psychotropic drugs in general? *Lancet Psychiatry* 2, 756-764, doi:10.1016/S2215-0366(15)00214-X (2015).

19   van Os, J. "Schizophrenia" does not exist. *Bmj* 352, i375, doi:10.1136/bmj.i375 (2016).

20   Aho, K. Medicalizing mental health: a phenomenological alternative. *J Med Humanit* 29, 243-259, doi:10.1007/s10912-008-9065-1 (2008).

21   Pattyn, E., Verhaeghe, M., Sercu, C. & Bracke, P. Medicalizing versus psychologizing mental illness: what are the implications for help seeking and stigma? A general population study. *Soc Psychiatry Psychiatr Epidemiol* 48, 1637-1645, doi:10.1007/s00127-013-0671-5 (2013).

22   Pickersgill, M. D. Debating DSM-5: diagnosis and the sociology of critique. *J Med Ethics* 40, 521-525, doi:10.1136/medethics-2013-101762 (2014).

23   Croxtall, J. D. Clonidine extended-release: in attention-deficit hyperactivity disorder. *Paediatr Drugs* 13, 329-336, doi:10.2165/11208100-000000000-00000 (2011).

24   Singh, I. & Rose, N. Biomarkers in psychiatry. *Nature* 460, 202-207, doi:10.1038/460202a (2009).

25   Correll, C. U., Kishimoto, T. & Kane, J. M. Randomized controlled trials in schizophrenia: opportunities, limitations, and trial design alternatives. *Dialogues Clin Neurosci* 13, 155-172 (2011).

26   Fleischhacker, W. W. *et al.* Placebo or active control trials of antipsychotic drugs? *Arch Gen Psychiatry* 60, 458-464, doi:10.1001/archpsyc.60.5.458 (2003).

27   Kemmler, G., Hummer, M., Widschwendter, C. & Fleischhacker, W. W. Dropout rates in placebo-controlled and active-control clinical trials of antipsychotic drugs: a meta-analysis. *Arch Gen Psychiatry* 62, 1305-1312, doi:10.1001/archpsyc.62.12.1305 (2005).

28   Miner, J. & Hoffhines, A. The discovery of aspirin's antithrombotic effects. *Tex Heart Inst J* 34, 179-186 (2007).

29   Vane, J. R. & Botting, R. M. The mechanism of action of aspirin. *Thromb Res* 110, 255-258 (2003).

30   Zaman, M. A., Oparil, S. & Calhoun, D. A. Drugs targeting the renin-angiotensin-aldosterone system. *Nat Rev Drug Discov* 1, 621-636, doi:10.1038/nrd873 (2002).

31   De Clercq, E. Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *Int J Antimicrob Agents* 33, 307-320, doi:10.1016/j.ijantimicag.2008.10.010 (2009).

32   Sheiner, L. B. Learning versus confirming in clinical drug development. *Clin Pharmacol Ther* 61, 275-291, doi:10.1016/S0009-9236(97)90160-0 (1997).

33   European Commission. (European Commission, Brussels, Belgium, 2014).

34   Cochrane, A. L. *Effectiveness and efficiency: random reflections on health services*. (Nuffield Provincial Hospitals Trust, 1972).

35   Insel, T. R. *et al.* Innovative solutions to novel drug development in mental health. *Neurosci Biobehav Rev* 37, 2438-2444, doi:10.1016/j.neubiorev.2013.03.022 (2013).

36   Morris, R. G. *et al.* Consensus Statement on European Brain Research: the need to expand brain research in Europe - 2015. *Eur J Neurosci* 44, 1919-1926, doi:10.1111/ejn.13236 (2016).

37   Nutt, D. & Goodwin, G. ECNP Summit on the future of CNS drug research in Europe 2011: report prepared for ECNP by David Nutt and Guy Goodwin. *Eur Neuropsychopharmacol* 21, 495-499, doi:10.1016/j.euroneuro.2011.05.004 (2011).

4

38   Rising, K., Bacchetti, P. & Bero, L. Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. *PLoS Med* 5, e217; discussion e217, doi:10.1371/journal.pmed.0050217 (2008).

39   The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, U.S. Department of Health Education and Welfare & Office of the Secretary. (U.S. Federal Government, Washington DC, USA, 1979).

40   Bacon, F. Franc. Baconis de Verulamio / Summi Angliae Cancellarii /Novum organum scientiarum. [Francis Bacon of St. Albans Lord Chancellor of England. A 'New Instrument' for the sciences] Lugd. Bat: apud Adrianum Wiingaerde, et Franciscum Moiardum. Aphorism XLVI (pages 45-46). (1645).

41   McMahon, A. D., Conway, D. I., Macdonald, T. M. & McInnes, G. T. The unintended consequences of clinical trials regulations. *PLoS Med* 3, e1000131, doi:10.1371/journal.pmed.1000131 (2009).

42   MoreTrials Campaign. Updated open Letter to EMA & ICH: From 5 research organisations and an international consortium of 119 health researchers in 22 countries. http://moretrials.net/the-problem/2016. Accessed 16 May 2017.

43   Phipps, A. I. *et al.* KRAS-mutation status in relation to colorectal cancer survival: the joint impact of correlated tumour markers. *Br J Cancer* 108, 1757-1764, doi:10.1038/bjc.2013.118 (2013).

44   Andreyev, H. J., Norman, A. R., Cunningham, D., Oates, J. R. & Clarke, P. A. Kirsten ras mutations in patients with colorectal cancer: the multicenter "RASCAL" study. *J Natl Cancer Inst* 90, 675-684 (1998).

45   Roila, F. & Cortesi, E. Quality of life as a primary end point in oncology. *Ann Oncol* 12 Suppl 3, S3-6 (2001).

46   Revicki, D. A., Kleinman, L. & Cella, D. A history of health-related quality of life outcomes in psychiatry. *Dialogues in clinical neuroscience* 16, 127-135 (2014).

47   Schandelmaier, S. *et al.* Planning and reporting of quality-of-life outcomes in cancer trials. *Ann Oncol* 26, 1966-1973, doi:10.1093/annonc/mdv283 (2015).

48   Jacobsen, P. B., Davis, K. & Cella, D. Assessing quality of life in research and clinical practice. *Oncology (Williston Park)* 16, 133-139 (2002).

49   van Leeuwen, M. *et al.* Recruiting long-term survivors of European Organisation for Research and Treatment of Cancer phase III clinical trials into quality of life studies: challenges and opportunities. *Eur J Cancer* 50, 1957-1963, doi:10.1016/j.ejca.2014.04.018 (2014).

Chapter 4.2

Outcomes of EMA marketing authorization applications: does partnering have an influence?

C.A. van den Bogert; P.C. Souverein; M. Putzeist, PhD, Susan Janssen; G.H. Koëter; H.G.M. Leufkens

ABSTRACT

Objectives
New medicines in the end-stage of development that fail marketing authorization application (MAA) cause huge financial losses and missed therapeutic opportunities. The effect of acquisition on MAA failure rate is unknown. Therefore, we studied whether the acquisition of new active substances (NAS) was associated with the outcome of MAA.

Methods
We identified the originators of all NAS evaluated by the EMA from 2009 until 2013 by a systematic crosscheck of online resources. Each NAS was categorized either as self-originated or acquired. The outcome was a positive or a negative MAA. Characteristics of MAAs were quantified and compared between approved and failed MAAs. Furthermore, we stratified by company size of the originator and of the applicant.

Results
Hundred seventy-two NAS dossiers were included, of which 103 were self-originated and 69 were acquired. Overall, acquired NAS and self-originated NAS had approximately the same MAA failure percentages (23.2% and 22.3%, respectively). NAS acquired as sole products had the highest failure percentage (38%). NAS acquired by small companies had a higher failure percentage (47%) than NAS acquired by medium or large companies (17%).

Conclusions
Acquired NAS are equally successful in obtaining marketing authorization as self-originated NAS. A small company size was significantly associated with a negative outcome of MAA, especially among NAS acquired by small companies. Small companies may benefit from a more extensive use of the available EMA services in their efforts of bringing new products to the European market.

INTRODUCTION

Combining strengths in drug discovery and development by partnering and outsourcing has been shown to increase the likelihood of success in bringing new products to the clinic[1]. In this context, a recent analysis by Lincker et al. from the European Medicines Agency (EMA) investigated the origin of new active substances (NASs) that were approved to enter the European market between 2010 and 2012[2]. They showed that a substantial proportion (~45%) of the 94 NASs approved in this time originated from small companies, academic institutions and public–private partnerships, and that most had been transferred to larger companies during development. However, this analysis did not assess whether the origin of a NAS is also associated with approval success or failure. Therefore, in this study we investigated whether self-originated NASs differed from acquired NASs with regard to the outcome of marketing authorization applications (MAAs) to the EMA.

METHODS

The study had a cross-sectional design. All data were collected from publicly available, online sources. We included all centralized market authorization application evaluations by the Committee for Medicinal Products for Human Use (CHMP) from 2009 to 2013 (both positive and negative) for analysis. Only products meeting the NAS definition were included in the dataset. An overview of all new active substances (NAS) submitted for marketing authorization can be found in the European Medicines Agency annual reports of the relevant calendar years. To identify the originator of the NAS, we searched the European Public Assessment Report (EPAR) on the EMA website, the archives of Scrip Intelligence (Scrip usually announces acquisitions, mergers and license agreements), PubMed (for the notified sponsors of the published first-in-human trials of the product) and Google (for other online pharmaceutical journal news items). Furthermore, we checked the website of the companies involved with each NAS in our dataset.

NAS were categorized according to their acquisition history in two groups, using complementary definitions of acquired and self-originated. Acquisition was defined as licensed, purchased or otherwise acquired from outside the sponsoring company (for example, from another company, a university, a government agency, or an individual). A NAS was self-originated if it was developed as a result of clinical research conducted either entirely by the sponsor of the MAA,

by a wholly owned subsidiary of the originating company or by an entity that is wholly a part of the organization of the MAA sponsor. The CMR International Pharmaceutical R&D Factbook also used this definition, kindly provided on request by Thomson Reuters.

We further characterized the acquisition as whole product acquisition, company merger/acquisition or partial license agreement. The decision algorithm for characterizing the acquisition started by deciding whether the company was acquired (or merged), or whether only the product was acquired. If an acquisition or a merger was the case, the acquisition was characterized as a company merger/acquisition. Then, we assessed whether the acquiring company obtained the unlimited, global marketing license and development rights of the substance, or whether obtained only a limited marketing license or development rights (e.g. for a specific geographic area). The unlimited, global acquisitions were classified as whole product acquisitions, and the limited acquisitions were classified as partial license agreements. As an example, Johnson and Johnson gained the global rights for carisbamate from its originator SK Holdings. Therefore, we classified the acquisition of carisbamate as a whole product acquisition.

The dependent outcome variable was MAA approval or failure, calculated as the percentage of NAS that failed. Approval was defined as a positive opinion from the CHMP regarding the marketing authorization in the EU. Failure was defined as either a negative opinion from the CHMP, or a withdrawal of the application by the sponsor before the CHMP issued an opinion.

We categorized the company size of the marketing authorization applicant on a three-level scale, according to the Scrip Pharmaceutical League Table on revenue (top-150) of the corresponding calendar year. If a company was ranked in the top 20 of the table, it was categorized as large, if it was between position 21 and 150 as medium, and if it was not listed in the table as small. Other studies investigating marketing authorization outcomes also used this approach[3,4] We categorized the company size of the originator using the same approach. The originator was defined as the company or organization that discovered the molecule up to the clinical stage of development. Key characteristics of the two acquisition groups were described, including the therapeutic area (based on ICD-10 classification), whether the scientific advice procedure was used by the sponsor, orphan drug status, type of the NAS (large-molecule biological or a small-molecule new chemical entity) and whether the NAS was previously approved elsewhere. If the same active substance was submitted more than once, we checked the first reviewed substance for previous approval outside the EU and classified the later reviewed substance(s) as previously approved elsewhere. Publicly available information in the EMA CHMP reports from 2009-2013 provided these data[5-7].

We quantified the trends of acquisitions and compared the failure percentages among acquired versus self-originated NAS over the consecutive calendar years. Furthermore, we tested whether the failure percentage of self-originated NAS was different from the failure percentage of the acquired NAS.

RESULTS

The CHMP issued an opinion on 172 NASs in the period 2009–2013 under the centralized MAA procedure. Of these 172, 133 (77%) were granted a positive opinion. Overall, 54 (41%) of these approved products originated from small companies, non-governmental organizations or academic institutions (figure 4.2.1), again highlighting the importance of these sources of new products for the European market. Also in line with previous research[4], NASs from small applicants had a higher failure rate (40%) than did those from medium- or large-sized applicants (17%; figure 4.2.2). Univariate χ^2 statistics identified no difference between the self-originated and the acquired groups ($\chi^2 = 0.02$; p = 1.000). We stratified the failure and approval percentages by the company size of the marketing authorization applicant, as this variable was independently associated with approval or failure ($\chi^2 = 10.04$; p = 0.002).
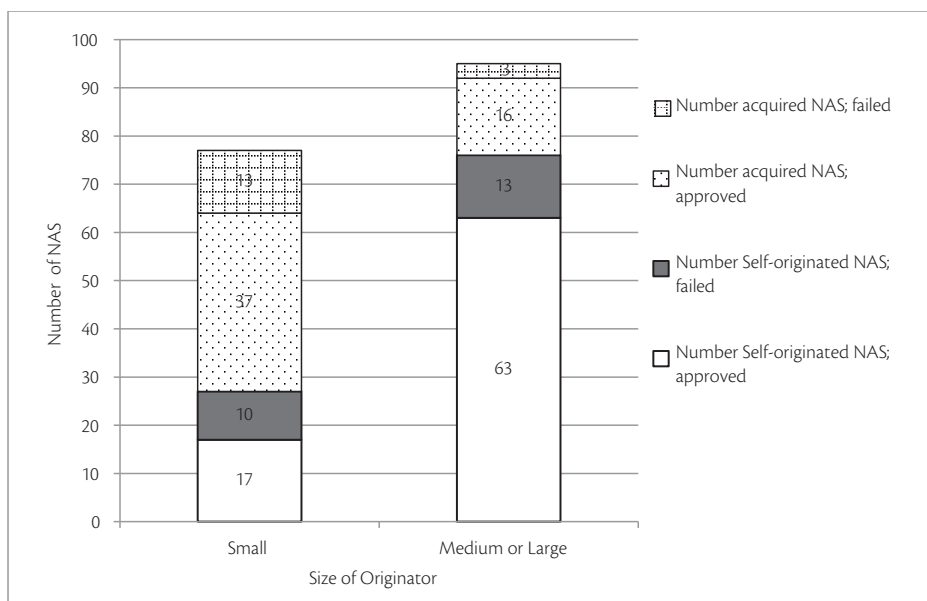
FIGURE 4.2.1 Number of New Active Substances approved and failed, stratified by company size of originator.
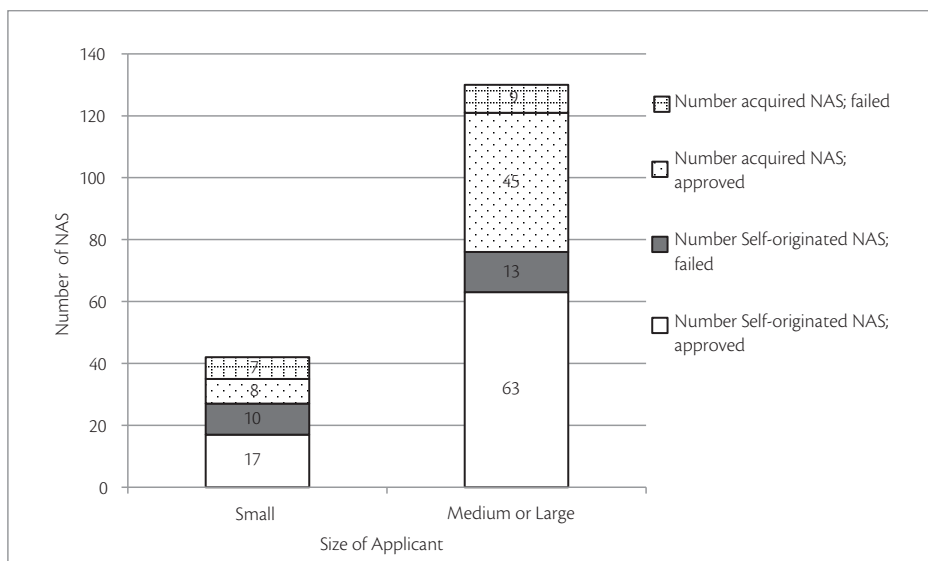
FIGURE 4.2.2 Number of New Active Substances approved and failed, stratified by company size of applicant.

Overall, 69 (40%) of the investigated NASs were acquired (table 4.2.1). We found no difference in the overall percentage of MAA failures between acquired and self-originated NASs (23% and 22%, respectively. When stratifying acquisition by the size of the applicant as depicted in figure 4.2.2, NASs that were self-originated by small applicants had a lower percentage of MAA failures (37%) than did NASs acquired by small applicants (47%). In the medium-/large-sized applicant group, we found no difference between the percentages of failed MAAs for acquired NASs and for self-originated NASs (both 17%; figure 4.2.2).

Considering the 69 product acquisitions in more detail (figure 4.2.3), most acquisitions (26) were whole product acquisitions (that is, where the acquirer took over responsibility for all further development and marketing), followed by whole company acquisitions (22) and partial license agreements (21) (for example, allowing the acquirer to market the product in a specific geographic region). Notably, when comparing failure percentages among the different types of acquisition, the failure rate of whole product acquisitions (38%) was twice the failure rate of products with partial license agreements (19%) and four times the failure rate (9%) of whole company acquisitions (Supplementary information S2). One possible explanation for the lower failure rates in the second two groups is because the development team in those situations remains the same or at least the new developers continue to have access to their expertise.

TABLE 4.2.1 Characteristics of NAS, submitted to the EMA for MAA from 2009 to 2013.

Characteristics of NAS	Total (n = 172)	NAS is self-originated (n = 103)	NAS is acquired (n = 69)
	N (% of 172)	N (% of 103)	N (% of 69)
Company size originator			
Small	77 (44.8%)	27 (26.2%)	50 (72.5%)
Medium/large	95 (55.2%)	76 (73.8%)	19 (27.5%)
Company Size applicant			
Small	42 (24.4%)	27 (26.2%)	15 (21.7%)
Medium/large	130 (75.6%)	76 (73.8%)	54 (78.3%)
Disease area			
Infectious diseases (including vaccines)	31 (18.0%)	15 (14.6%)	16 (23.2%)
Neoplasms	44 (25.6%)	26 (25.2%)	18 (26.1%)
Hematological diseases	8 (4.7%)	7 (6.8%)	1 (1.4%)
Neurological and behavioral diseases	18 (10.5%)	8 (7.8%)	10 (14.5%)
Cardiovascular and respiratory diseases	19 (11.0%)	14 (13.6%)	5 (7.2%)
Rheumatology and other musculoskeletal diseases	12 (7.0%)	11 (10.7%)	1 (1.4%)
Endocrine diseases	17 (9.9%)	9 (8.7%)	8 (11.6%)
Other	23 (13.4%)	13 (12.6%)	10 (14.5%)
Scientific advice received			
Yes	130 (75.6%)	76 (73.8%)	54 (78.3%)
No	42 (24.4%)	27 (26.2%)	15 (21.7%)
Orphan Drug status			
Orphan Drug	39 (22.7%)	25 (24.3%)	14 (20.3%)
No Orphan Drug	133 (77.3%)	78 (75.7%)	55 (79.7%)
Product type*			
NCE	106 (61.6%)	56 (54.4%)	50 (72.5%)
Biological, radiopharmaceutical or other ATMP	66 (38.4%)	47 (45.6%)	19 (27.5%)
NAS previously approved elsewhere			
Yes	64 (37.2%)	34 (33.0%)	30 (43.5%)
No	108 (62.8%)	69 (67.0%)	39 (56.5%)

* χ^2 = 5.7, p = 0.025. Abbreviations: NAS, New Active Substance; MAA, Marketing Authorization Application; ICD, International Classification of Diseases; NCE, New Chemical Entity; ATMP, Advanced Therapeutic Medical Product.

4
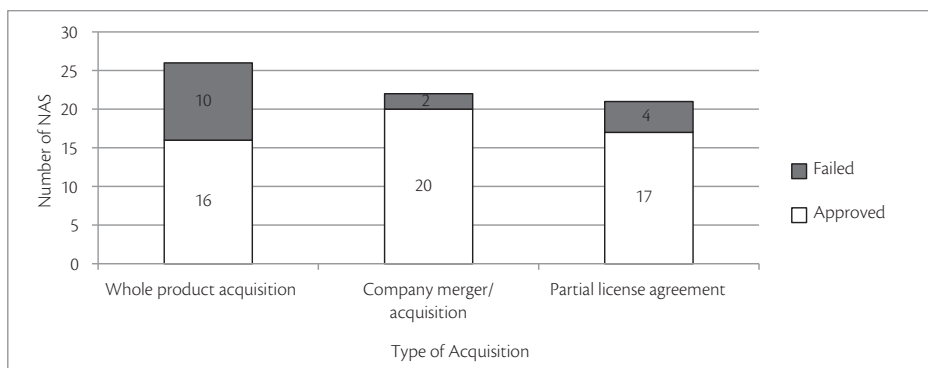
FIGURE 4.2.3 Number of acquired new active substances (NAS) with positive and negative marketing authorization application outcomes, categorized by the type of acquisition. Total number of acquired NAS = 69.

DISCUSSION

In our study, four out of ten of the new products reviewed by the EMA were not discovered by the company that applied initially for marketing authorization, but acquired when entering or proceeding in clinical development. Our results show that these products from an external source were equally successful in obtaining a positive MAA outcome as self-originated products. This finding is surprising, as one would expect companies buy only the good and promising products in development to strengthen their pipeline and to increase their revenue. The company size of the marketing authorization applicant was the only variable with a strong association with the MAA outcome. Apparently, small companies had more difficulty in bringing products to the European market, despite services offered to them by the EMA through the small- and medium-sized enterprise (SME)-office.

The CHMP bases its opinion with regard to the marketing authorization on the evidence of clinical effect, the size and relevance of this effect and the quality of the product[8]. In their scientific evaluation, regulators will not consider whether the submitting company acquired or originated a medicine. However, according to the scientific literature, failure of pharmaceutical products is associated with various characteristics of the product, size of the submitting company and aspects of the development. Related factors include company size[3,4], orphan drug designation[4], compliance with the scientific advice of the EMA[4], a good development plan[8,9], target novelty[10,11], selection of outcomes in the pivotal studies[12] and the type of the product[13-15]. Our findings show that these factors fluctuated heavily over the five calendar years, implying there is neither a structural nor a causal relationship between these characteristics and the MAA outcome. The only vari-

able that showed a consistent association with the MAA outcome over the past five calendar years was the size of the marketing authorization applicant.

Looking at our findings, it is unclear why small companies acquire NAS given the unfavorable MAA success. Furthermore, our findings indicate that more NAS originated by a small company that were subsequently acquired by another company obtained a positive MAA as compared when the NAS was submitted by the originating small company itself. Thus, performing the whole development process may not always be within the capability of small originators. Licensing out the late-phase development or the regulatory submission to another pharmaceutical company is in those cases a good way of bringing new products to the market. A company may for example not be sufficiently equipped or experienced for conducting the large clinical pivotal studies required for marketing authorization. Having no office in an EU member state may be another reason for licensing out the European marketing rights. Furthermore, especially the large pharmaceutical industry uses acquisition to strengthen their pipeline. From a business perspective, it is recommendable to have more than just one lead asset in the pipeline. Literature shows that the acquisition of products is an effective portfolio strategy according to innovation business models[16-18]. Creating development utilities and employing the right experts is costly. Once this infrastructure is in place, it is more effective to scale-up the development activities performed by the same people and equipment.

Interestingly, the NAS that were acquired as a sole product had a higher failure percentage than the NAS that were acquired by company mergers. In the case of a company merger, the same development team stays responsible for the process, whereas when a sole product is acquired, the transition of people working with the product may lead to distraction. In case of a license agreement, the success may depend for a large part on the communication between the collaborating companies, when two development teams work on the product in usually separate parts of the world.

We may have missed some acquisitions, namely those that we could not identify in the searched media sources. As a result, some products may be wrongfully classified as self-originated. However, selective misclassification with regard to the outcome is unlikely and hence we believe misclassification influenced our results to little extent.

Following up on the findings of the paper by Lincker et al.[2], our study adds some important insights. First, it confirmed that small companies remain an important source of the new products developed for the European market. From 2009 until 2013, 54 (41%) of the approved products came from small companies, NGOs or universities. The finding that small companies are still facing higher

failure percentages at the time of MAA submission, both for acquired as for self-originated NAS, illustrates the dialogue between these small companies and regulators needs improvement. Thirdly, our findings imply that the acquisitions performed by small companies were not always the best choices, as they faced a high failure percentage among their acquired products. The data quantitatively illustrate that pharmaceutical innovation has to come for an important part from small research and biotech companies, and the medium- and large-sized pharmaceutical industry is better in turning such innovative products into marketable drug products.

In conclusion, the finding that small companies still have higher failure percentages for MAAs — both for acquired and self-originated NASs — than do larger companies implies that the regulatory dialogue between small companies and regulators needs improvement. Furthermore, our findings indicate that small companies may not be optimally skilled for in licensing, as the NASs acquired by small applicants had a particularly high percentage (47%) of failed MAAs. Overall, as also indicated by previous analyses, collaborating with larger companies may be a more effective strategy for smaller companies to bring new products successfully to the European market.

REFERENCES

1. DiMasi JA, Feldman L, Seckler A, Wilson A. Trends in risks associated with new drug development: success rates for investigational drugs. *Clin Pharmacol Ther.* Mar 2010;87(3):272-277.

2. Lincker H, Ziogas C, Carr M, Porta N, Eichler HG. Regulatory watch: Where do new medicines originate from in the EU? *Nat Rev Drug Discov.* Feb 2014;13(2):92-93.

3. Putzeist M, Mantel-Teeuwisse AK, Gispen-De Wied CC, Hoes AW, Leufkens HG. Regulatory scientific advice in drug development: does company size make a difference? *Eur J Clin Pharmacol.* Feb 2011;67(2):157-164.

4. Regnstrom J, Koenig F, Aronsson B, et al. Factors associated with success of market authorisation applications for pharmaceutical drugs submitted to the European Medicines Agency. *Eur J Clin Pharmacol.* Jan 2010;66(1):39-48.

5. European Medicines Agency. Performance of the Agency's scientific procedures: 2011 report for medicinal products for human use. London, UK: EMA.

6. European Medicines Agency. Annual report 2012 Annex 21 - Performance of the centralised procedure (human medicines). London, UK: EMA.

7. European Medicines Agency. Annual report 2013 Annex 9. CHMP opinions in 2013 on medicinal products for human use. London, UK: EMA; 2014.

8. Putzeist M, Mantel-Teeuwisse AK, Aronsson B, et al. Factors influencing non-approval of new drugs in Europe. *Nat Rev Drug Discov.* Dec 2012;11(12):903-904.

9. Eichler HG, Aronsson B, Abadie E, Salmonson T. New drug approval success rate in Europe in 2009. *Nat Rev Drug Discov.* May 2010;9(5):355-356.

10. Ringel M, Tollman P, Hersch G, Schulze U. Does size matter in R&D productivity? If not, what does? *Nat Rev Drug Discov.* Dec 2013;12(12):901-902.

11. Walker I, Newell H. Do molecularly targeted agents in oncology have reduced attrition rates? *Nat Rev Drug Discov.* Jan 2009;8(1):15-16.

12. Hartmann M, Mayer-Nicolai C, Pfaff O. Approval probabilities and regulatory review patterns for anticancer drugs in the European Union. *Crit Rev Oncol Hematol.* Feb 19 2013.

13. Pignatti F, Aronsson B, Gate N, et al. The review of drug applications submitted to the European Medicines Evaluation Agency: frequently raised objections, and outcome. *Eur J Clin Pharmacol.* Dec 2002;58(9):573-580.

14. Heemstra HE, de Vrueh RL, van Weely S, Buller HA, Leufkens HG. Predictors of orphan drug approval in the European Union. *Eur J Clin Pharmacol.* May 2008;64(5):545-552.

15. Putzeist M, Heemstra HE, Garcia JL, et al. Determinants for successful marketing authorisation of orphan medicinal products in the EU. *Drug Discov Today.* Apr 2012;17(7-8):352-358.

16. Girotra K, Terwiesch C, Ulrich KT. Valuing R&D projects in a portfolio: Evidence from the pharmaceutical industry. *Manage Sci.* Sep 2007;53(9):1452-1466.

17. Hoetker G, Agarwal R. Death hurts, but it isn't fatal: The postexit diffusion of knowledge created by innovative companies. *Acad Manage J.* Apr 2007;50(2):446-467.

18. Johnson G. What place for R&D in tomorrow's drug industry? *Drug Discov Today.* Mar 1996;1(3):117-121.

Chapter 5

Risk indicator taxonomy for supervision of clinical trials on medicinal products

Authors:
Peter M.J.M. Jongen*; Cornelis A. van den Bogert*; Catharina W.E. van de Laar; Kim Notenboom; Elysée T.M. Hille; Ingrid Hegger

* Contributed equally

ABSTRACT

Objectives

To facilitate a risk-based approach for the supervision of clinical trials on medicinal products, we identified and categorized indicators that may present an elevated safety and/or ethical risk for participants, and/or for data integrity. The indicators are relevant for all stakeholders including participants, regulatory bodies, health care inspectorates, sponsors and trial sites.

Methods

The sources of indicators included Medline (using the search terms risk-based/-triggered/-driven oversight/monitoring/inspection), relevant documents from websites of regulatory authorities in Europe, North America and Australia, and results of a brainstorm session organized for experts working in the field. Indicators were classified according to risk area (safety and ethical, data integrity, or both).

Results

In total, we identified 69 risk indicators that were categorized in six branch-levels of the taxonomy. We visualized the taxonomy in a tree-structure to clearly distinguish individual indicators. In addition to readily detectable risk indicators, more context-related aspects determine the final impact of the trial and constitute further components in risk assessment. Context-related aspects include potential high media attention, consequences for the reputation of medical research, and the socioeconomic situation in the geographic region and have to be considered on a case-by-case basis.

Conclusions

We identified a wide array of risk indicators for clinical trials on medicinal products and we used a tree-structure to incorporate the indicators identified to clearly distinguish individual indicators and to enable efficient use of the indicators. The overview of indicators may facilitate multiple stakeholders in developing structured risk assessment (identification and analysis) for supervising clinical trials on medicinal products. Stakeholders can interpret and prioritize the indicators from their own perspective.

INTRODUCTION

Protecting subject safety and data integrity is fundamental in the supervision of clinical trials on medicinal products, as emphasized by the first statement in the preamble of the European clinical trial regulation: "In a clinical trial the rights, safety, dignity and well-being of subjects should be protected and the data generated should be robust"[1]. For those involved in design, management and supervision of clinical trials, it is relevant to discern where and when these principles are 'at risk'. To be able to do so, awareness of risk indicators is essential.

Risks are present in clinical trials in all stages, and the risks vary depending on the perspective and the stakeholder [2]. Regulatory bodies may adopt risk-based approaches to carry out their supervisory tasks effectively, such as targeting and conducting Good Clinical Practice (GCP) inspections and the assessment of clinical trial application. In addition, the European Clinical Trial Regulation encourages sponsors to adopt risk-adapted approaches in monitoring practices and data management of trials. Over the past years, regulators publically shared their viewpoints on risk-based approaches for regulatory activities with respect to clinical trials on medicinal products in guidance documents and in presentations, postulating a variety of indicators for risk assessment [3-5]. In addition, other stakeholders shared their thoughts and opinions on risk-adapted approaches for clinical trials[6-8]. The focus of the risk indicators may depend on the perspective. For example, guidance documents from regulatory agencies dealing with marketing authorizations focus on data integrity, ethics committees focus on the safety of trial subjects whereas sponsors might have additional interest related to potential cost savings. A comprehensive risk assessment should therefore consider risk indicators proposed by a diversity of sources.

Inspectorates can use risk indicators to identify high-risk trials and to prioritize and balance risks in their inspection policy decisions. Institutional and national review boards can use these indicators in their risk- assessment. In addition, sponsors and trial sites can use such indicators to be more aware of eventual risks before and during the trial, and avoid mistakes or failure.

An overview and categorization of indicators was proposed by Hartmann and colleagues[9], based on documents from both industry and regulators[5,10]. Two examples of indicators to trigger extra supervision such as a GCP-inspection are a start-up company conducting the trial and a vulnerable population participating in the trial. A new company conducting a trial can be an indicator of an elevated risk for data integrity, because the company has no experience in recording clinical data. The severity of the disease of the subjects studied in a trial is an indicator of potential elevated patient safety risk.

5

For the purpose of a risk-based approach of GCP-inspections, we previously developed a set of risk indicators for the supervision of clinical trials. By applying this set in the development of risk-based supervision, we felt the need for a systematic, comprehensive set of risk indicators. This article describes how we identified risk indicators and incorporated them in a tree-structure to clearly distinguish individual indicators and to enable efficient use of the indicators. The indicator taxonomy described is intended to provide a basis for structured risk assessment and prioritization. The stakeholders who use the indicator taxonomy can subsequently assign priorities or weights to the indicators relative to other indicators.

METHODS

The definition of risk by ISO guideline 31000:2009 is 'effect of uncertainty on objectives'[2,11]. We translated 'effect of uncertainty' to the context of clinical trials as 'uncertain events with a negative effect'. The general objective of clinical trials is to answer one of more specific research questions with valid and accurate data (data integrity), without compromising the safety and ethics of participants. Hence, we define risk in clinical trials as 'uncertain events that have a negative effect on participant safety or ethics and/or on data integrity'.

Uncertain events can be characterized by the likelihood that the event occurs and the severity if the event occurs[12,13]. Examples of uncertain events are serious adverse events such as adverse drug reactions, enrolling participants without informed consent, and breaches of the protocol.

Risk indicators can facilitate the detection of trials with an increased risk for participants and/or data integrity. We defined risk indicator as properties of a trial that are readily detectable and presenting a safety or ethical risk for trial participants and/or a risk for data integrity. We collected risk indicators from existing literature, guideline documents, and results of an open brainstorm session in March 2014. Expert regulators from the National Institute of Public Health and the Environment, the Healthcare Inspectorate, the Central Committee on Research Involving Human Subjects, and the Medicines Evaluation Board participated in the brainstorm session. The proceedings of the session were the starting point for the collection of risk indicators. We subsequently searched the websites of regulatory authorities in Europe, North America and Australia for official guidelines on risk-based oversight of clinical trials mentioning indicators for risk assessment. Furthermore, we searched in Medline for scientific publications. We used a Boolean search strategy with terms including risk-based/-triggered/-driven oversight/monitoring, in combination with 'clinical

trials', and 'factors', or 'indicators'. We screened the abstracts, retrieved full text publications, and derived indicators intended for risk assessment described in these publications.

Next, we developed a taxonomy based on the collected indicators. We categorized the indicators according to the aspect of the trial covered, i.e. 'What', 'By whom', and 'How'. The branch 'What' contains indicators related to the investigational medicinal products involved in the trial. The branch 'By whom' contains indicators related to the clinical trial site, institute and sponsor conducting the trial. Finally, the branch 'How' contains indicators related to the design and actual conduct of the trial. We allocated the indicators to one of the three main categories, and structured them using a tree-structure.

RESULTS

The summarizing scheme of indicators is shown in figure 5.1, and the full taxonomy is included in an interactive supplementary file available online. We will further explain the indicators in the text below. In all figures, we underscored indicators if they mainly affect the safety and/or ethical risk for participants to distinguish between these and indicators mainly affecting the risk for data integrity. The indicators in bold font were considered to equally affect the risk for participants and for data integrity. To be able to keep track of the indicators, we coded the indicators alphabetical for the first level, and a number for each next level. We based the sequence and codes solely on their relation to the taxonomy and not on their importance or priority.

"What": investigational medicinal product (figure 5.2)

A1. Knowledge of Investigational Medicinal Product (IMP) in humans

First in man phase I trials and efficacy-exploring phase II trials (trials with IMPs not yet authorized for marketing) pose an increased safety risk for participants because any new chemical entity may have unknown/not anticipated pharmacological/toxicological effects that can be harmful. For an authorized IMP, the likelihood of safety events can be increased if the IMP is tested for a new indication, in a new subset of patients, or in a new formulation because the IMP may influence the disease pathophysiology or other medicines leading to unexpected safety issues[5,6,9,14,15].
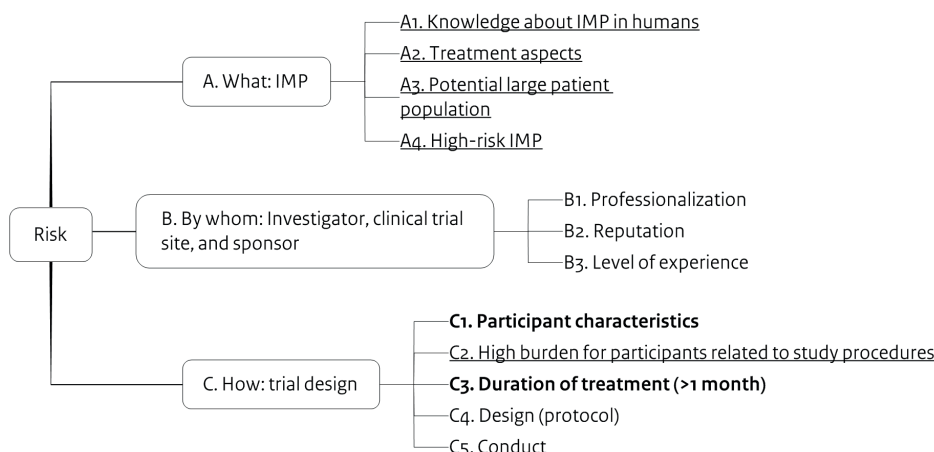
FIGURE 5.1 First two levels of tree-structured taxonomy. IMP = investigational medicinal product. Non-underscored indicators in non-bold font mainly affect the risk for data integrity. Underscored indicators mainly affect the safety and/or ethical risk for participants. Bold indicators were considered to equally affect the risk for participants and for data integrity.

A2. Treatment aspects

Treatment aspects of the IMP can cause safety events in a clinical trial. If there are no treatment options available for the expected side effects, participants are exposed to a higher safety risk compared to expected side effects for which an effective treatment exists. If potential side effects affect the vital systems, the safety risk may also be increased[15].
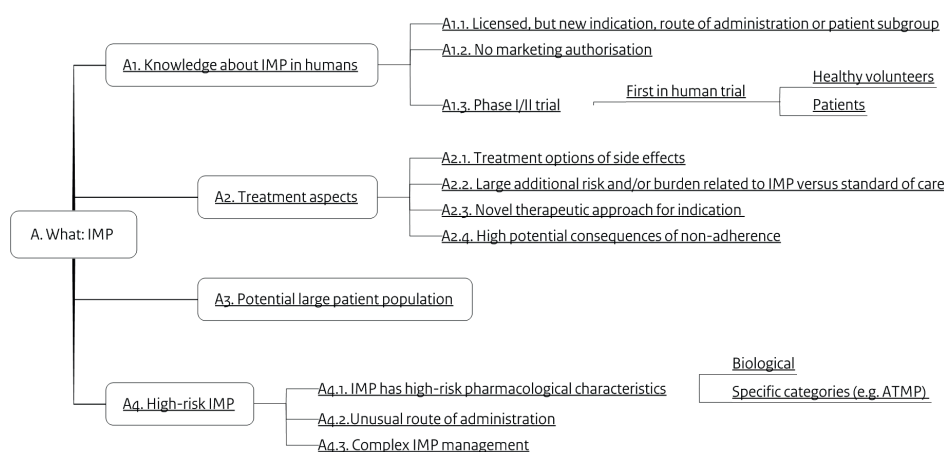
FIGURE 5.2 Indicators related to "what": the investigational medicinal product (IMP). All indicators in the what-category were considered to mainly affect the safety and/or ethical risk for participants.

In some trials, the likelihood of safety issues related to the trial treatment is much higher compared to the treatment that the trial participants would have

received as standard of care. This is the case, for example, in clinical trials on new oncology products with terminal cancer patients. These patients would normally receive relatively safe palliative treatment, such as analgesics. By participating in trials to investigate new oncolytic IMPs, they are exposed to potentially unsafe experimental treatments without benefit for themselves. Such trials have thus an elevated safety risk, compared to trials in which the treatment-related safety risk is similar to the standard of care. If a trial investigates a novel therapeutic approach for the target indication, this also presents a safety risk. This is the case if a disease is treated in a trial using a different type of product or procedure. Another aspect that may influence the safety risk is the severity of non-adherence consequences, i.e. discontinuation of a treatment regimen by the participant that can lead to deterioration of health[5,6,9,14].

A3. Potential large patient population
If the IMP is indicated for a disease with a potential large patient population, this presents an increased risk for data integrity. After marketing approval, the consequences of incorrectly performed studies may harm a potential large number of patients. This indicator is therefore of particular interest for authorities involved in marketing authorization approval[16].

A4. High-risk IMP
The IMP involved in a trial can hold characteristics that increase the safety risk for participants. The three types identified of risk-increasing characteristics are the pharmacology, route of administration, and stability.

Pharmacological aspects of the IMP can increase the likelihood of serious adverse events during the trial, such as a less well known mechanism of action (biological, advanced therapeutic IMP), nonlinear kinetics, a narrow therapeutic window, and the severity of non-efficacy consequences (e.g. with antibiotics). This likelihood may be also increased if for the IMP an unusual formulation or a new route of administration is used. In addition, the safety risk can be increased if the IMP has a low stability. For example, some IMPs are prone to manufacturing and storage errors also including reconstitution aspects. This can be a challenge in a trial setting[5-7,9,14,15].

"By whom": Investigator, clinical trial site and sponsor (figure 5.3)
B1. Professionalization
Sponsors linked to established, commercial organizations have usually more resources for taking appropriate measures to ensure high professional standards for data integrity [9]. Compared to non-commercial, non-profit, and start-up or-

224 Chapter 5

ganizations, the commercial organizations are thus expected to have a higher degree of professionalism and therefore present a lower risk for data integrity[16].

B2. Reputation

A good record of accomplishment of an investigator, trial site or sponsor as established from previous trials may provide reassurance on their performance in current and future trials. Lacking an inspection history[10], or negative findings during previous inspections, calls for further information on whether data integrity is appropriately ensured. The reputation of an organization or investigator will be negatively affected by whistle blower reports, inconsistencies in regulatory documents, and previous rejections of clinical trial applications by the medical research ethics committee (MREC).

When a supervisor needs to assess clinical trials results from regions beyond its authority, the strength of the supervisor in charge can also influence the risk for data integrity[16].
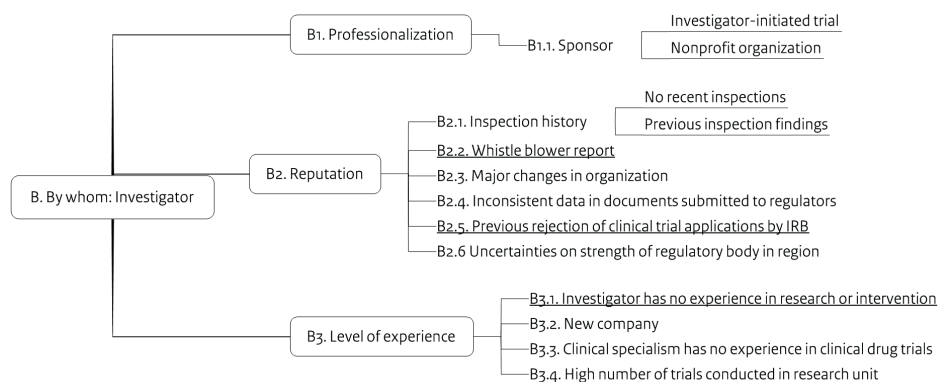
FIGURE 5.3 Lower-level indicators related to "by whom": the sponsors, investigators, and clinical research site. IRB = Institutional Review Board. Non-underscored indicators in non-bold font mainly affect the risk for data integrity. Underscored indicators mainly affect the safety and/or ethical risk for participants.

B3. Level of experience

Clinical investigators for whom the type of intervention in the trial is less familiar, can overlook safety signals more easily, leading to an increased safety risk[4].

Companies new to the field of clinical research have potentially fewer standardized procedures and qualified staff to promote working according to the guidelines. Hence, an increased data integrity risk is perceived for trials conducted by unexperienced companies[4,9]. The same can be the case for investigators from clinical specialisms that have no experience in clinical trials[4,5].

If a clinical trial site or research unit conducts many clinical trials, they are probably more familiar with appropriate procedures, thus presenting a lower risk in particular for data integrity. However, if there are issues at organizational level, more participants will be affected and the impact of the issue will be larger[16].

"How": Trial design and conduct (figure 5.4)

C1. Participant characteristics

The number and type of trial participants can influence the safety and ethical risk for participants, and the risk for data integrity. On the one hand, if a trial enrolls a large number or participants, the severity of systematic mistakes in following the protocol is increased because more participants will be affected. On the other hand, the likelihood of mistakes could be increased in trials with low enrollment (for example, in rare indications), because the investigators gain less experience with the protocol[3,10,14].

Phase 3 trials have usually a large sample size, so phase 3 can be an indicator for increased safety risk because of a large number of participants. More importantly however, these trials are often pivotal in marketing authorization applications and the consequences of data integrity breaches in phase 3 trials may affect many patients in the future[17].

From an ethical perspective, trials including participants unable to give informed consent present a higher ethical risk because the decision to participate will be made not by the participant, but by its legal representative. Not only the complexity of such a procedure increases the likelihood of data integrity breaches but also the societal impact of incidents in such trials will be larger compared to trials including participants who decided themselves whether or not to enroll[5,6].

More specifically from a safety perspective, trials enrolling vulnerable patients may have an increased risk. This can be age-related (in particular children or elderly), or disease related (in particular psychiatric and somatic patients may be prone to safety issues). In addition, the safety risk may be increased in pregnant or breast-feeding women[3,4,6,7,14].

When a trial enrolls participants in an emergency setting, the risk for data integrity is increased. Given the urgent situation and time-pressure, trial procedures may not be followed according to the protocol[3,6].

C2. High burden for participants related to study procedures

The burden imposed by a trial on a subject is an important contributor to the safety and ethical risk for trial participants[3,5-7,14,18]. Ethical concerns may arise if

participants themselves can expect little or no benefit from participating in the trial (e.g. early-phase trials of unauthorized IMPs on healthy subjects).

If the trial procedures resemble the standard-of-care-procedures for the patient, the likelihood of unexpected safety issues will be low. Nevertheless, consideration should also be given to the additional psychological or physical burden imposed by the trial on a patient when compared to the standard of care. In addition to the IMP (indicator A2.3, figure 5.2), such burden may arise from procedures such as invasive physical examinations, time-consuming diagnostic procedures, extra hospitalizations, and radioactive labelling. For example, if biopsies are collected in a trial for further biomarker or genetic investigation, the safety risk may be increased because participants would not undergo a biopsy as standard of care.

C3. Duration of treatment

If participants undergo the trial treatment for a longer time, the safety risk for participants and the risk for data integrity increase[16]. Generally, a longer exposure may increase the possible occurrence of adverse drug reactions compared to a short exposure[5]. In addition, the chance of non-adherence with trial treatment and of participant withdrawal will be higher in trials with a long duration of treatment.

C4. Design

The risk for data integrity may also relate to design aspects of a trial. Indicators were classified under 'design' if they are laid down a priori in the protocol. If the trial is conducted at more than one trial site, trial oversight by the sponsor can be more difficult and, therefore the likelihood of data integrity issues is increased[4,19].

Trials consisting of multiple stages (e.g. a randomized and a single-arm extension stage) present a higher risk because the procedures for data management are prone to mistakes during transition between trial stages[16].

If several organizations are involved in trial procedures, data integrity issues may occur more likely than when everything is done by a single stakeholder[19]. If many stakeholders are responsible for trial procedures, data management and communication is more challenging than if all procedures are managed by one stakeholder. For example, if different contract research organizations (CROs) are appointed with tasks, miscommunication can arise. Moreover, if tasks change among the stakeholders, the data integrity can be compromised due to a different interpretation of the protocol[17].

A more thorough analysis of trial methodology revealed indicators to identify trials allowing for a more flexible interpretation of the results can increase the data integrity risk. First, the use of subjective, or soft, endpoints allows for more

intra-investigator and inter-investigator variation or bias in interpretation of data. Secondly, new assessment tools may give variances in measurement that was not detected during validation. Thirdly, an unnecessarily complex design such as multiple treatment arms, placebo formulations or data management systems increases the likelihood of protocol breaches. Last, a non-inferiority design incorporates more room for subjectivity and bias compared to a superiority design[5,17].
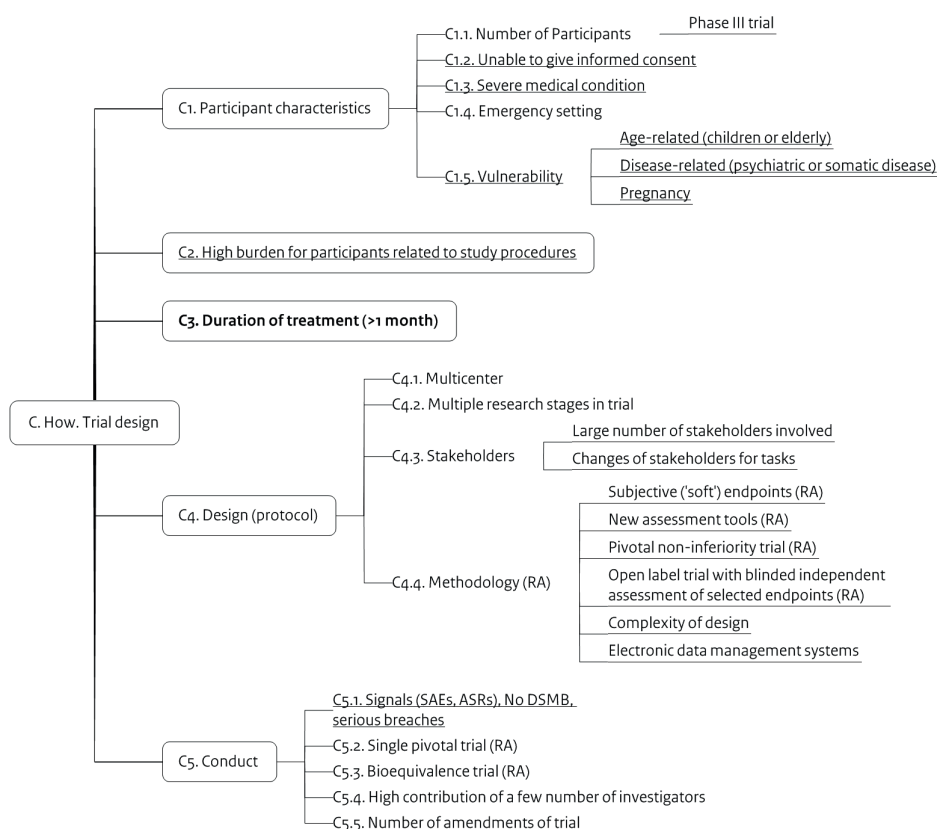

FIGURE 5.4 Lower-level indicators related to "how": the clinical trial design and conduct. RA = Regulatory Authority; SAE = Serious Adverse Event; ASR = Annual Safety Report; DSMB= Data Safety Monitoring Board. Non-underscored indicators in non-bold font mainly affect the risk for data integrity. Underscored indicators mainly affect the safety and/ or ethical risk for participants. Bold indicators were considered to equally affect the risk for participants and for data integrity.

C5. Conduct

Several indicators are only evident after the trial has started and cannot be identified upfront. The safety and ethical risk (and also the risk for data integrity) can be elevated if many safety signals (serious adverse event reports, annual

safety report) have already been recorded in the course of the trial[16]. In addition, the safety risk is increased if no Data Monitoring Board is appointed to monitor if any issues emerge. Serious breaches with the protocol are also an indicator of an increased safety risk[5,6].

Data integrity can be at risk if there is only a single pivotal trial conducted for the authorization of a new IMP. Any data integrity issues can in that case directly misinform the authorities, physicians and patients, in their evaluation on the safety, efficacy and quality of an IMP. A bioequivalence trial can also elevate the risk for data integrity because bioequivalence trials usually collect less extensive, more falsifiable data than confirmatory trials of novel products[20].

If there is a high contribution of few investigators to the conduct and data analysis of the trial, the likelihood is increased that intentional or non-intentional mistakes remain unnoticed[17].

The amount of amendments can also be an indicator of increased risk for data integrity. In case of many amendments, the original design of the trial was apparently of low quality because many modifications were needed. The change of procedures may then lead to a distraction from the original objectives of the trial and the relevance of the trial data[17].

DISCUSSION

We have identified a wide array of clinical trial characteristics that may contribute to the safety and ethical risk for participants and to the risk for data integrity in clinical trials. These risk indicators are presented in a systematic manner using a tree-structure. The taxonomy as proposed is based on current thinking about and development of risk-based supervision as presented in scientific publications, presentations and guidance documents. Moreover, the taxonomy concisely describes indicators relevant across all stakeholders (i.e. participants, inspectorates, regulatory authorities, trial sites, sponsors and target populations[2]).

The tree-structure may show the relevancy of an individual indicator with respect to the nature of the risk, the relationship between indicators and it may reveal new indicators for particular stakeholders. We highlighted indicators with a direct relationship to safety and ethical risk for participants (e.g. first-in-human trials, vulnerable trial subjects) or to data integrity (e.g. pivotal trials). Nonetheless, the usability of an individual risk indicator will also rely on ease of evaluation and on accessibility of information. Indicators such as the phase of a trial, the license status of the IMP or the number of trial participants' are

self-explanatory and easy to evaluate while several design and conduct aspects require assessments that are more thorough and access to detailed trial information.

In the process of risk management, the indicators facilitate structured risk assessment, and more specifically, risk identification and analysis. The taxonomy can be used as checklist as a starting point for stakeholders to identify factors that impact the risk of clinical trials. Moreover, the taxonomy facilitates risk analysis because the indicators and their explanation provide further understanding of the risks, their causes and their consequences. Examples of assessment procedures of different stakeholders include monitoring procedures by trial sponsors[6-8,21], identification of triggers for GCP-inspections by regulators[3-5,10], and ethical review by MRECs[5,22]. In the development of risk ranking tools, the stakeholder's perspective will determine the interpretation and prioritization of the indicators. Regulatory bodies may wish to assign risk ranks to trials in order to define risk-based requirements for clinical trials or to select trials for inspection. For example, from the perspective of GCP inspectorates, data integrity indicators such as the absence of previous inspection reports will receive a high priority, whereas this may be of less priority for the MREC.

The translation of structured risk assessment to practical decision-making, for example in supervision activities, needs also consideration of the context-related aspects that may influence the societal impact of the trial. The context-related aspects of the trial include but are not limited to potential high media attention, consequences for reputation of medical research and the situation in the geographic region where the clinical trial is conducted. These aspects are time-depended and place-depended and are difficult to quantify or capture in objective indicators. Context related aspects may arise on a case-by-case basis and are identified based on experience.

Risk assessment of clinical trials is a complex process that will likely benefit from a systematic and structured approach. Indicators related to the IMP, the investigators involved and the trial design provide building blocks for an assessment procedure. However, an overly strict interpretation of a risk assessment based on indicator scoring while leaving no room for experienced based intuition is a potential pitfall. No evidence exists that categorizing clinical trials according to strict criteria leads to improved practices when it replaces intuitive assessment. Such intuitive assessment may for example link information from several individual risk indicators and flag risk increases not easily revealed by theoretical models[23-25]. Nevertheless, this caveat is no reason to set aside systematic risk assessment approaches. It merely suggests to leave room for professional judgement of the results of the risk assessment within a decision making process[26].

In addition to stakeholders involved in supervision, (potential) trial participants have to make their personal risk assessment leading to the decision whether or not to participate in and complete a trial. However, current guidelines and existing literature on risk assessment of clinical trials do not specifically address the risk assessment to be made by a potential trial participant. Therefore, the perspective of participants in risk assessment of clinical trials may be underrepresented in the taxonomy. Aspects that are likely important for trial participants to make their personal assessment include insurance, financial compensation, safety of trial facilities, how they are informed by the recruiting investigator, and access to a trial when no other treatments are effective (in particular in the case of life-threatening diseases).

In conclusion, the risk indicator taxonomy presented here can contribute to risk-based supervision of clinical trials. The taxonomy can facilitate a structured identification and analysis of risk for participants and for data integrity. Besides the risk indicators, a case-by-case assessment of the context remains indispensable. We aimed to cover the diversity of risk indicators in an exhaustive manner but since on an ongoing basis more experience is gained in management and supervision of clinical trials, others will likely add new and other viewpoints to this subject. To our knowledge, our taxonomy contains the most comprehensive overview of risk indicators currently available in the literature and may therefore provide a valuable template for further research on and development of risk-based supervision of clinical trials.

ACKNOWLEDGEMENTS

REFERENCES

1. European Commission. Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC.: European Commission; 2014.

2. European Clinical Research Infrastructures Network. Guideline on risk management for clinical research. Paris, France: ECRIN; 2015.

3. European Medicines Agency. Reflection paper on risk based quality management in clinical trials. London, UK: EMA; 2013.

4. Food and Drug Administration Center for Drug Evaluation and Research. Guidance for Industry. Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring. Silver Spring, Maryland USA: FDA; 2013.

5. Medicines and Healthcare Products Regulatory Agency. Risk-adapted Approaches to the Management of Clinical Trials of Investigational Medicinal Products. London, UK: MHRA; 2011.

6. Brosteanu O, Houben P, Ihrig K, et al. Risk analysis and risk adapted on-site monitoring in noncommercial clinical trials. *Clinical trials.* Dec 2009;6(6):585-596.

7. Journot V, Pignon JP, Gaultier C, et al. Validation of a risk-assessment scale and a risk-adapted monitoring plan for academic clinical research studies--the Pre-Optimon study. *Contemporary clinical trials.* Jan 2011;32(1):16-24.

8. Morrison BW, Cochran CJ, White JG, et al. Monitoring the quality of conduct of clinical trials: a survey of current practices. *Clinical trials.* Jun 2011;8(3):342-349.

9. Hartmann M, Hartmann-Vareilles F. Concepts for the Risk-Based Regulation of Clinical Research on Medicines and Medical Devices. *Drug Information Journal.* 2012;46(5):545-554.

10. Health Canada Health Products and Food Branch Inspectorate. Inspection Strategy for Clinical Trials. Ottawa, Canada: Health Canada; 2002.

11. European Medicines Agency. EMA recommends suspension of medicines over flawed studies at Semler Research Centre. London, UK: EMA; 2016.

12. European Medicines Agency. EMA recommends suspension of medicines due to unreliable studies from Micro Therapeutic Research Labs. London, UK: EMA; 2017.

13. Imran M, Najmi AK, Rashid MF, Tabrez S, Shah MA. Clinical research regulation in India-history, development, initiatives, challenges and controversies: Still long way to go. *J Pharm Bioallied Sci.* Jan 2013;5(1):2-9.

14. Netherlands Federation of University Medical Centres. Kwaliteitsborging mensgebonden onderzoek 2.0 (Dutch) 2012. Houten, the Netherlands: NFU; 2012.

15. Kenter MJ, Cohen AF. Establishing risk of human experimentation with drugs: lessons from TGN1412. *Lancet.* Oct 14 2006;368(9544):1387-1391.

16. Clinical trial risk model brainstorm session organised by Healthcare Inspectorate and National Institute of Public Health and the Environment. Date of session: 14 March 2014. Utrecht, the Netherlands.

17. European Medicines Agency. Points to consider for assessors, inspectors and EMA inspection coordinators on the identification of triggers for the selection of applications for "routine" and/or "for cause" inspections, their investigation and scope of such inspections. London, UK: EMA; 2013.

5

18. Rid A, Emanuel EJ, Wendler D. Evaluating the risks of clinical research. *Jama.* Oct 6 2010;304(13):1472-1479.

19. Organisation for Economic Co-operation and Development Global Science Forum. Facilitating International Cooperation in Non-Commercial Clinical Trials. Paris, France: OECD Publishing; 2011.

20. European Medicines Agency. Guidance on triggers for inspections of bioequivalence trials. London, UK: EMA; 2013.

21. Tudur Smith C, Williamson P, Jones A, Smyth A, Hewer SL, Gamble C. Risk-proportionate clinical trial monitoring: an example approach from a non-commercial trials unit. *Trials.* 2014;15:127.

22. Rid A. Setting risk thresholds in biomedical research: lessons from the debate about minimal risk. *Monash Bioeth Rev.* Mar-Jun 2014;32(1-2):63-85.

23. Dobrow MJ, Goel V, Upshur RE. Evidence-based health policy: context and utilisation. *Social science & medicine.* Jan 2004;58(1):207-217.

24. Strydom WF, Funke N, Nienaber S, Nortje K, Steyn M. Evidence-based policymaking: A review. *South African Jounal of Science.* June 9 2010;106(5/6):Art.#249, 248 pages.

25. Dobrow MJ, Goel V, Lemieux-Charles L, Black NA. The impact of context on evidence utilization: a framework for expert groups developing health policy recommendations. *Social science & medicine.* Oct 2006;63(7):1811-1824.

26. Cevallos M, Zullig S, Christen A, et al. A pilot test of the new Swiss regulatory procedure for categorizing clinical trials by risk: A randomized controlled trial. *Clinical trials.* Jun 29 2015.

Chapter 6

General discussion

GENERAL FINDINGS

This thesis addressed two central research questions. First, what is the fate of clinical drug trials, and what are determinants of failure to reach the optimal fate; and, secondly, how routinely registered data during the clinical trial application and Institutional Review Board (IRB) review process can be used for risk-based approaches in clinical trial governance. This discussion chapter describes the main results, reflects on methodological strengths and limitations, provides recommendations for further research and practice, leaves room for personal reflections and gives an overall conclusion.

We sought an answer to the first research question by a series of analyses of a cohort consisting of all 622 clinical drug trials reviewed by the Dutch accredited IRBs in 2007. In the evaluation of the fate of these trials, we used a framework in which we distinguish between the scientific fate and the development fate. We used this framework to handle the two worlds in which drug trials are conducted: the world of science and the world of drug development. On the one hand, a clinical drug trial is an academic exercise, aiming to answer a research question on a micro level, and to contribute to the body of scientific knowledge on the macro level[1-3]. On the other hand, clinical drug trials are conducted to bring new products to patients [4]. Although these two worlds are closely intertwined, the judgement of trials diverges depending on which world's perspective is chosen. From a scientific perspective, a trial investigating a relevant question about one or more drugs, using the appropriate design and methods, recruited the sample size needed, and is published in the academic literature, can be judged as a scientific success, regardless the results. From a drug development perspective, however, this trial would be a failure if the results are non-significant, as the drug of interest unlikely ends up as treatment against the studied disease, hence failing to deliver therapeutic profits to patients and financial profits to companies and investors[5-7]. Thus, using the term 'trial failure' to describe trials that demonstrated non-superiority or inferiority of the drug of interest over the comparator arm has nothing to do with science, but reflects the development objectives of trials. The involved scientists may also be inclined to view such trials as failed, as they are probably conducting the trial in the hope of developing new therapies for patients in need. Moreover, negative results will probably result in less attention to their work, and will unlikely be published in the most prestigious journals.

In the introduction, it was described that the scientific fate of a trial is in the first place determined by its methodology. Of particular interest are the methods used in first in man (FIM) trials. A persistent tendency to conduct these trials in a one-size-fits-all approach, which means basing dose escalation on the maximum

tolerated dose[8]. It has been clearly demonstrated and argued that pharmacology (pharmacokinetics and pharmacodynamics) should be the main targets for these early-phase drug trials[9]. This means that pharmacokinetics (PK) and pharmacodynamics (PD) should inform dose calculation and escalation. We investigated in chapter 2.1 the design choices of first in man trials (FIM) and found that one in three FIM trials calculates the FIM dose using only the no observed adverse effect level (NOAEL) from the preclinical experiments. Furthermore, one in five FIM trials only used allometric scales to extrapolate the preclinical information to the FIM dose.

When combining the completion and reporting rates, 283 trials were completed or discontinued in a justified way and published in the scientific literature, with no protocol-publication discrepancies in the primary endpoint. Thus, as illustrated in figure 6.1, the scientific fate of 45% of the trials that were reviewed by the Dutch IRBs in 2007 is that they have transparently contributed to the scientific body of knowledge with regard to their primary endpoints. This cumulative finding from chapter 2.2, 2.3 and 2.4 is in general consistent with the literature[10-15], although the protocol-publication discrepancies in primary endpoints was remarkably lower in our cohort compared to previous research[16,17]. Should we consequently declare the remaining 55% (342 trials) of the 2007 cohort as wasted and unethical? Ten trials must at least be omitted from this percentage, as these were still running at the moment of closing the data collection. For the 38 trials that were rejected by the IRBs and those that were approved, but never started recruitment, the waste remains limited to the time devoted to the preparation of the protocol and the IRB-review. Because no participants were included, these trials have no ethical consequences. The research waste is more substantial among the 291 trials (47% of the cohort) that started recruiting participants, but discontinued for questionable reasons and/or were selectively reported. Moreover, the scientific fate of unpublished trials is undesirable from an ethical perspective, as this fate means that the burdens (e.g. potentially risky experimental treatments, body fluid and/or tissue collections) are not contributing to an answer to the research question, nor to the body of knowledge, that is available for investigators embarking on a future trial, doctors, meta-analysts, clinical guideline committees, and patients.

In line with findings by Van Lent et al.[18,19], investigator-initiated trials were at increased risk of discontinuation due to recruitment failure and of protocol-publication discrepancies in the primary endpoints. The sponsor type was not associated with non-publication. Publications of industry-funded trials were more likely to have a positive direction of conclusion compared to investigator-initiated trials: 73% of the industry-sponsored trials had a positive direction of conclusion

with regard to the drug of interest, versus 50% of the 14 investigator-initiated trials with industry (co-)funding and 60% of the 40 investigator-initiated trials with no industry (co-)funding. This finding is consistent with the literature[20,21], and can be explained by the larger sample size of industry-sponsored trials. Indeed, the median included sample size of the 150 published industry-sponsored trials in our cohort was 316.5 participants (interquartile range 72.3-801.0), versus a median included sample size of 32.0 (interquartile range 19.0-72.5) for investigator-initiated trials (co-)funded by the industry and a median of 24.0 (interquartile range 16.0-67.5) for investigator-initiated trials with no industry (co-)funding. It is much easier to obtain statistical significance in trials with around 300 participants compared to trials with only 30 participants. Another explanation is that a larger proportion of industry-sponsored trials is building on previous findings and has therefore a higher probability of success. We could not identify the magnitude or determinants of publication bias, as we did not know the direction of findings or conclusions of the unpublished trials.
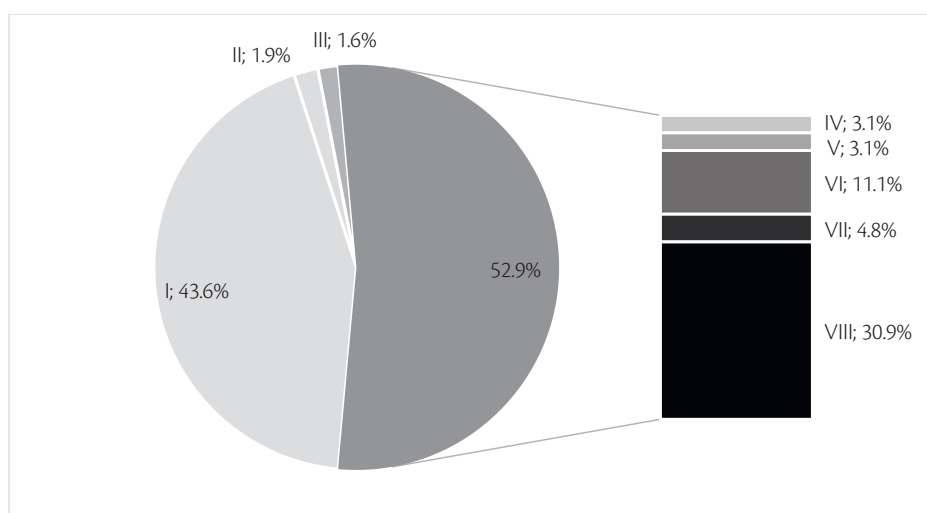
FIGURE 6.1 Scientific fate of the 622 clinical drug trials reviewed by the Dutch accredited Institutional Review Boards in 2007. Legend: I = Completed and published with primary endpoints according to protocol; II = Justified discontinued and published with primary endpoints according to protocol; III = Still running at end of study period; IV = Rejected by the IRB; V = Approved by the IRB, never started; VI = Unjustified discontinued; VII = Published with protocol-publication discrepancy primary endpoint; VIII = Unpublished.

In chapter 3, the mechanisms of selective reporting, the largest contributor to an unfavorable scientific fate, were explored in more detail. Using a literature-based determinant taxonomy, we proposed a causal scheme of how selective reporting putatively works. Our theory distinguishes between necessary causes (e.g. motivations and specific means) and component causes (effect modifiers),

and between causes at the level of the individual and causes at the level of the system and wider environment. This also shows the complex etiology of the scientific fate of drug trials – often, many of these determinants co-occur and can influence the outcome.

The fate of the 2007 cohort from the perspective of drug development was investigated in chapter 2.5. We demonstrated that 35% of the trials were included in the regulatory product dossier of the drug marketing authority in the EU. The majority of non-included trials were not included because marketing authorization was not (yet) sought for the drug product (33% of the cohort), or because they were neither designed nor required for submission to the marketing authorization authorities (27% of the cohort). Thus, one-third of the drug trials in the cohort were initiated by the company that was developing or marketing the drug product, but were not part of the regulatory product dossier 9 years after IRB-review. Using submission to the regulatory product dossier as a proxy for the development fate of drug trials, it can be concluded that the development fate of one-third of the drug trials was unsuccessful 9 years after IRB-review. This percentage could still improve, however, as some development programs might not have been completed yet by the end of our follow-up. Therefore, some trials could still be included in future submissions to the regulatory authorities.

Success in clinical drug development was further investigated in chapter 4, where we contrasted the phase 3 clinical drug trial landscape of the EU of oncology with psychiatry (chapter 4.1) and looked at success versus failure of initial marketing authorization applications of new drugs to the European Medicines Agency (EMA) (chapter 4.2). The contribution of chapter 4 to the research questions of this thesis is to provide a broader view on drug development. The disease area (chapter 4.1) and business strategy (chapter 4.2) is a major determinant of success once a drug enters the clinical development phase, in particular for small companies. Given that oncology has been a thriving success market over the past years, many small and large companies as well as academic groups are focusing on this area. This is reflected by the large share of oncology in the total phase 3 clinical trial activity of the EU. In other therapeutic areas, such as psychiatry, progress in understanding disease biology and development of new therapies has been very limited, causing a logical downward spiral: if a disease (area) has a track record of development failures, commercial investors will be reluctant to invest in new R&D projects and medical faculties might shift research priorities as well. Governments have recognized the need for translational research in resource-deprived areas such as psychiatric disorders and antimicrobial resistance and launched extensive publically funded research programs[22,23]. However, a declining pharmaceutical productivity is not necessarily curable by investing

more money [24]. Drug development has in the current situation arrived at a cru-cial point, as the notion gains support that the return on investments in R&D are insufficient[4]. Investors and company managers in the pharmaceutical sector have started to realize that downsizing R&D is an efficient method to reduce costs without losing revenue[24]. Whether this is true or not, no new therapies will be developed without new research. Finding ways to stimulate and incentiv-ize innovative research is therefore likely one of the main challenges in drug development of this time.

RISK-BASED GOVERNANCE

In chapter 5, risk-based supervision of clinical trials was analyzed by creating a taxonomy of risk indicators mentioned in literature and guidelines. All indicators were of qualitative origin: they were not identified in empirical studies, but in guidelines and viewpoints published by regulatory agencies and academics. We found no evidence that if a quantitative model based on risk indicators is used to prioritize high-risk trials for trial supervision would lead to the identification of more deficiencies compared to other approaches. Furthermore, we identified a trade-off: trials with higher consequences of eventual Good Clinical Practice (GCP) issues have in general a lower likelihood of GCP issues. Trials where eventual issues with safety or data integrity (the two pillars of GCP) are severe will be surrounded by a more adequate quality system, and be conducted by more experienced investigators, compared to trials where such issues do little harm to participant safety and/or date integrity. Thus, trial supervisors may choose on the one hand to visit trials where they will probably find a number of issues, but where these issues have limited actual impact. On the other hand, they can choose to spend their time with visiting high-risk trials (where potential issues would have serious impact), but likely find few issues.

According to chapter 5, the evaluation of risks in clinical trials can be ap-proached systematically by using risk indicators. Our indicator taxonomy can serve as starting point for such risk evaluation. However, to use the indicator taxonomy to rank trials based on their risk, we also found that it is the decision of the supervisor which indicator deserves the heaviest weight. The finding that both phase 1 and phase 3 are indicators illustrates this. For some supervisors, phase 1 trials will have the highest risk (for safety), whereas other supervisors consider phase 3 as more risky (for data integrity). Which indicators to prioritize relates to the function or mandate of the specific supervisor. These decisions are guided by experience, consensus and intuition, also taking the context of

the trial into account. Some trials that have few risk-elevating indicators may still need to be assigned as high risk, for example because the topic has a high potential for media attention. These risks are not directly related to GCP, but were mentioned by supervisors as important factors in their decision-making. We found no evidence that a quantitative model can replace this.

METHODOLOGICAL REFLECTIONS

This thesis contains five chapters with studies addressing the two main research questions about the fate of trials and risk-based supervision. These studies have in common that they are all observational in nature; no experimental designs have been used, nor did we systematically summarize the results of other experiments. By having access to the national regulatory IRB database ToetsingOnline, we were able to use the best data source at our disposal, as this source contains all clinical trial applications and trial protocols submitted to IRBs in the Netherlands. In the study of (the fate of) trials, an IRB database is the source which is most complete with regard to selection of a cohort, as all clinical trials pass upon their inception through the portal of the IRB. Whether trials are registered or published in subsequent databases is uncertain – and does partially depend on the fate of the trial. Using, for example, clinicaltrials.gov to establish a cohort of clinical trials, will introduce a selection bias because not all trials are registered at this web portal.

Internal validity

In the studies of this thesis, we intended to provide insights in the occurrence and determinants of fate of drug trials. There are limitations – by design – in the interpretability of the determinants. These limitations originate in the non-experimental nature of the studies. Non-experimental studies only allow hypotheses generation, but not hypothesis testing and falsification. Based on the association of trial characteristics with the fate of trials, we have observed that trials with certain characteristics reached more often a certain fate, but we cannot demonstrate that characteristics cause the fate (or even play a role in it). The identified associations of non-publication with trial characteristics provide information on specific areas where, for example, non-publication is most prevalent, but do not elucidate why[25]. We cannot rule out that there is an underlying alternative explanation, other than chance, for the observed associations. Furthermore, the analyzed determinants are merely categorical trial

characteristics, which are useful as proxies and surrogate identifiers, but do not elucidate the clear-cut etiology of the investigated problems.

Notwithstanding these limitations of observational, inductive designs, we postulated in the chapters and in this discussion some causal inferences based on the observed associations. An advantage that we have in this process is that clinical trials were the units of analysis, and not, for example, humans or animals, who are immensely more complex objects to study. The immense complexity of human biology, combined with virtually infinite histories of exposures and experiences, and the fact that these factors may or may not be differential with regard to the variables under study, makes attempts to derive causal relations from associations obtained from observational research quite complicated. Although clinical trials are complex enterprises as well, the list of potential alternative causal explanations for observed associations is relatively limited. This practical assumption can serve as the basis for justification of a careful causal interpretation of the observed associations. It is also the only available option: the determinants that we studied cannot be allocated to trials at random.

In chapter 2, we analyzed several trial characteristics as potential determinants of the fate of trials. Information on these trial characteristics was obtained from the ABR-form[6]. A complicating issue in the analysis of the determinants was that many determinants were multicollinear. Of the 13 determinants that we investigated in chapter 2, we analyzed the 78 (13*12*0.5) possible single correlations between these determinants. Forty-nine of these single correlations were correlated ($p<0.05$) according to chi-square tests. The presence of multicollinearity becomes a problem in the interpretation of the association of the determinants with the outcome. For example, in chapter 2.3 we identified single center and phase 1 as two determinants of non-publication, two determinants that are also strongly correlated with each other (phase 1 trials are much more often single center compared to phase 2, 3 and 4 trials). So, remain single center trials more often unpublished because their correlation with the determinant phase 1, or is the correlation with non-publication also significant independently from the variable phase 1?

In chapter 2.3, we used multivariable logistic regression to assess the associations between the potential determinants and non-publication while adjusting for other determinants. In the other chapters, we did not perform logistic regression because there were too few outcomes (chapter 2.2 and 2.4), or because alternative reasons explained the results (chapter 2.5). However, even if there are sufficient

---

6   The clinical trial application form: the General Assessment and Registration (Dutch abbreviation: ABR)-form, mandatory and identical throughout the Netherlands.

outcomes to build a robust model and the parameters remain stable after adding collinear variables to the model; this does not completely resolve the problem of collinearity. Some trial subpopulations within the cohort can still yield a stable logistic regression model, while their interpretation may not be that straightforward. For example, phase 2 single center trials may have a lower likelihood of publication compared to phase 2 multicenter trials, whereas phase 3 single center trials may have the same likelihood of publication compared to phase 3 multicenter trials. Stratification would be the best method to dissolve these interactions and effect modifications. However, there were too many variables involved in the multicollinearity issue to stratify for all of them.

The correlations between characteristics are mostly quite logical. Using the example of the characteristics phase 1 and single center, phase 1 trials are in general conducted in a small number of often-healthy volunteers under fulltime observation in specialized phase 1 clinical research centers. Phase 2-, 3- and 4 trials require a larger sample size of patients that are not necessarily fulltime observed within the research clinic. These circumstantial differences explain why phase 1 trials are more often single center compared to other phase trials.

The trial characteristics that were investigated in chapter 2 as determinants can, as illustrated by figure 6.2, be subdivided into variables that are fixed upon the conception of the research question, and variables that are choices made by the trial investigator determined by methodological and practical possibilities. Thirdly, there were two legal status-variables that are added to the ABR-form because they have consequences in the determination whether the trial is permitted by the Dutch law (therapeutic effect expected), or in the requirements for the clinical trial application documents (drug approval status). Methodological and practical choices and the legal status variables are mostly determined by the set of characteristics that are fixed upon conception of the research question. The trial characteristics sponsor, phase, disease area, drug category, and participant category, are all more or less fixed upon conception of the research question. CRO involvement, centers, randomization, and prospective registration are characteristics defined by the practical choices that the investigators make during the operationalization of their research question.

This classification of determinants can help in the interpretation of the associations found in this thesis. Characteristics in the upper box of figure 6.2 may be upstream connected to the causal pathway of the fate of a trial, as they are defined together with the research question. Characteristics in the lower two boxes may be more or less logical consequences of the upper box or could be driven by other circumstances, and can be useful for policy makers aiming to improve the fate of trials. For example, single center trials were associated with

non-publication (chapter 2.3) and can be easily identified for supervision in order to improve their fate, regardless whether there is a causal connection between trials conducted at one study site and non-publication.
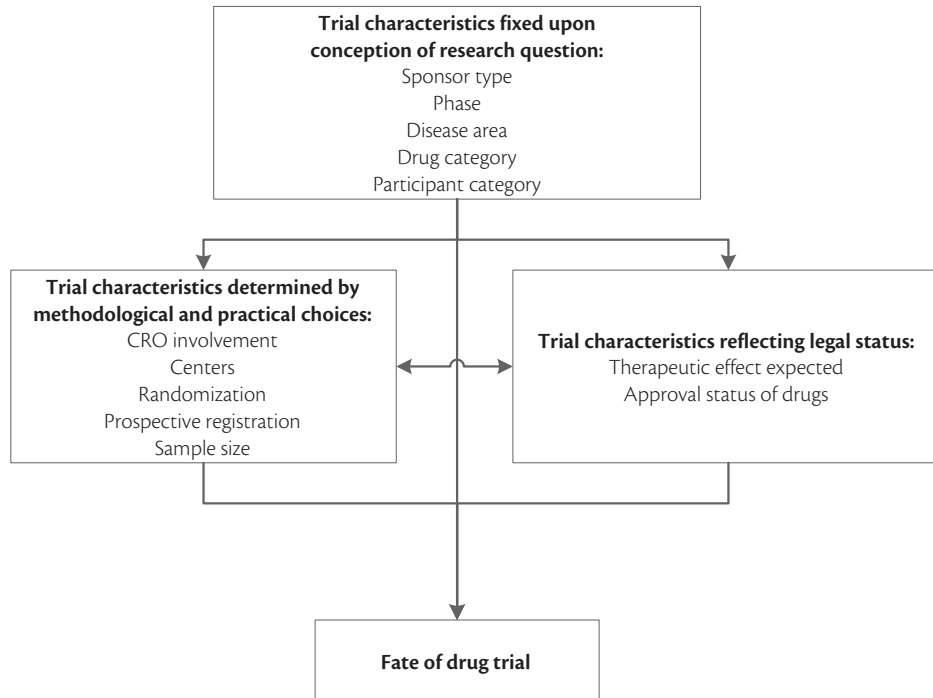

FIGURE 6.2 Categorization of trial characteristics

Some of the trial characteristics that we analyzed in chapter 2 were of limited value. The characteristics 'drug category' and 'participant category' had small numbers in at least one of the categories. An analysis of these characteristics as determinants of the trial fate would be, therefore, only be precise in much larger cohorts. Furthermore, 12 different disease areas remained after reduction and merging of some overlapping and/or small areas. Our sample size was too small to include these twelve disease areas in multivariable analysis. Because oncology is on some aspects different from other disease areas (in particular the early-phase trials), we included the disease area coded as oncology versus all other areas.

A potential determinant that we did not investigate was the IRB that reviewed the trial. Our data source provided us with data from 23 IRBs of widely different settings. Some IRBs mainly review phase 1 trials, whereas others are mostly reviewing investigator-initiated trials. Most IRBs are hospital-based, and the research topics may be very different per hospital, depending on circumstances

such as the available medical specialties, research groups, severity of illness of the patients and laboratory facilities. A number of these factors are associated with the outcomes of chapter 2. Thus, the association of the specific IRB with these outcomes would have been very difficult to interpret and could lead to un-intended and ill-founded conclusions. Moreover, a confidentiality agreement was signed for the research of this thesis, prohibiting publication of data that might identify individual trials, investigators, products, hospitals or companies. Thus, the IRB was not included as variable for methodological and practical reasons.

External validity

Because we selected the entire population of drug trials in the Netherlands meet-ing the inclusion criteria "drug trial" and "IRB-reviewed in 2007", the question of external validity concerns two aspects. The first aspect is whether our findings can be extrapolated to clinical drug trials conducted in the Netherlands in other years than 2007. The second aspect is whether the findings can be extrapolated to clinical drug trials not conducted in the Netherlands. Regarding the first aspect, appendix 2.2 concluded that the 2007 cohort was a representative reflection of the Dutch clinical trial landscape until 2015. Thus, the Dutch trial landscape has remained relatively constant in terms of the investigated trial characteristics.

Regarding the second aspect, it should be acknowledged that clinical drug research is currently extensively globalized[26,27]. Thus, our 'Dutch story' of chap-ter 2 tells only a very small part of the whole picture. In chapter 4, we tackled the issues using European data, which represents a larger geographical area, although extrapolation to other areas remains speculation. Nevertheless, the characteristics of the 2007-cohort compare well to cohort studies conducted in other countries[28-30], and the findings are generally within the range of these stud-ies. Moreover, studies investigating worldwide drug pipelines suggest a clinical trial landscape that is comparable to our analyses[31-33].

Increased attention has been paid to research waste and publication bias in popular books and journal articles[21], including the influential 2014 Lancet series on research waste[10-14]. Therefore, more awareness might exist on the importance of trials reaching a good scientific fate, causing (unbiased) publication percent-ages to increase. Two recent studies indeed suggest, for example, improved publication rates for phase 1 trials[34,35].

Database suitability

The centralized regulatory IRB database ToetsingOnline containing data about all clinical drug trials in the Netherlands (and of chapter 5, the EU) was the main data source for the studies of this thesis. As it was the first time that this

source was used for research, the exact opportunities and challenges were a priori unknown that would be encountered when working with this database. In an ideal situation, all data can be extracted from relational databases that – in this case – could be linked through a unique identifier, i.e. the IRB-number. The database did not contain the outcomes of interest of the studies in chapter 2.1-2.5. Several pieces of data needed to be manually retrieved, increasing the labor-intensiveness of our studies.

Another limitation was the many missing time registration fields in ToetsingOnline. Because of this missing information, we could only a non-optimal start-date of the trials in our "time-to-fate" analysis, as the only time registration in ToetsingOnline having no missing information was the date of IRB-approval. A more in-depth analysis of time intervals, for example between start and completion of trials, or between completion and publication, would have been informative, but resulted in a too substantial proportion of trials for which these data were unavailable, which would possibly have introduced a bias. In addition, because of missing data we had to make assumptions about whether some trials had started and were completed.

The type of research questions that are readily answerable based on the data from ToetsingOnline only can be used to broadly describe the trial landscape, to investigate what sort of trials are going on (such as done in Appendix 2.2 and chapter 6). This can be of use to regulators evaluating whether new types of trials are emerging and if they should train or hire expertise in certain areas, or adapt their policies to the changing clinical trial landscape. Moreover, the variables included in the ABR form can be used to facilitate systematic identification of risky trials, as some of the risk indicators can be identified in the ABR-form.

A barrier to using ToetsingOnline for research is its non-relational infrastructure. Whereas most databases have accessible functionalities to extract data, apply filters, etc., extractions from ToetsingOnline need to be performed through advanced programming by the IT-officer of the competent authority. ToetsingOnline lacks these functionalities because the database was not developed with the intention to monitor the fate of the trials on meta-level. Furthermore, clinical trials in the ToetsingOnline database are assigned a national clinical trial number. This number is currently not indexed by other databases such as Medline or public trial registries as standard practice, which does not allow for straightforward and quick identification of publications in connection to trial protocols (such as done in chapter 2.3 and 2.4).

IMPLICATIONS FOR FURTHER RESEARCH

The major knowledge gap left by our cohort study on the fate of drug trials for future research is the empirical identification of reasons why trials remain unpublished, and why publications have discrepancies with the protocol. Knowledge about reasons could help to implement specific solutions for the problem. To identify the reasons for non-publication, questionnaires were sent to all investigators, but principal investigators of only 55 out of the 240 unpublished trials (23%) responded. In the systematic review of chapter 3, we extracted determinants of selective reporting, which could be both reasons for selective reporting and characteristics of studies or investigators associated with selective reporting. Most of the reasons were extracted from opinion articles; therefore, more empirical data are needed to identify these. However, the actual methods to progress the knowledge about causal mechanisms will be a challenge.

It could be that investigators or sponsors are less interested in writing a scientific publication of a trial with no significant results or undesirable conclusions. For the sponsor, publishing negative results can mean a drop in sales, or the process of writing and publishing an article can be seen a waste of effort as the development fate of the trial is failed by the negative results. Similar incentives can be in place for investigators and editors to focus on publishing significant results. Moreover, in chapter 3, 11 other putative determinants were identified than the direction of the results for selective reporting practices, which may or may be not related to the direction of results or other personal prejudices. One of the determinants described in chapter 3 is, for example, the difficulty of publishing an academic peer-reviewed article. The possibility to report results in public registries would omit this hurdle. However, research shows that the reporting rates in these registries are still disappointing[36], suggesting that other motivations exist for selective reporting than the struggle to complete a peer-reviewed publication. So, as all reason is the slave of the passions[37], each camp can use evidence at its convenience. This problem will unlikely be resolved by conducting or systematically summarizing observational studies, as these are not suited for testing hypotheses[38]. All determinants mentioned in chapter 3 could play a role, and practical solutions need to address the putative causes as extensively as possible. Future observational research can be used to diagnose progress, and to identify areas with increased risk, but will unlikely resolve or disentangle the mechanism. As recent evidence shows that non-reporting and selective reporting is still prevalent[36,39], such monitoring remains relevant. This can become part of routine practice if the improvements to the data infrastructure of clinical trial databases are realized.

Another recommendation regards the development fate of clinical trials and drug products. In this thesis, we used marketing authorization as the desired development fate. However, although an important milestone, marketing authorization is just a step in the total lifecycle of a drug. From the perspective of the company, at the moment of marketing authorization, a drug has yet to deliver on its promise in terms of revenue. From the perspective of patients it is also unsure at the moment of marketing authorization what value the new drug will bring, whether they can afford it, what the long-term safety profile of the drug looks like, etcetera. Future research could, therefore, follow-up on the cohort of trials of chapter 2, and on the cohort of drugs of chapter 4.1, to investigate the fate of the products and trials on the long term. Research questions for such studies can include whether the drugs turned out as the intended generator of revenue for the developer; how the drugs performed in comparison to standard of care; whether there were (safety) issues downstream the lifecycle, and which determinants at the premarketing development stage were associated with these outcomes.

6

PRACTICAL IMPLICATIONS

Improving the fate of trials

The fate of the cohort of clinical drug trials investigated in this thesis was poor among a substantial proportion of the trials, regarding completion, reporting and complete reporting. The next question is then, how can these outcomes of fate be improved? This question entails two aspects: who (which stakeholder should improve?) and what (which practice or procedure should this stakeholder improve?). Looking at the stakeholders that play a role in clinical drug trials, there are, first, the stakeholders that actually conduct the trial: the investigators and sponsors. The latter are included here because they enable the conduct of trials in the first place. Secondly, legislators and regulators of country (and supranational) governments are involved in the regulation of the trials. In the Netherlands, all clinical drug trials have to be reviewed by IRBs that are accredited by the Central Committee on Research Involving Human Subjects (CCMO), of which the members are appointed by the Minister of Health, Welfare and Sports. Thus, in the Dutch system, local accredited IRBs, and in extension the CCMO, are the representatives of the government. The discussion below uses the Dutch situation as starting point and therefore, IRBs will be considered as the main regulatory stakeholder. Recommendations for IRBs may need to be extrapolated to other organizations mandated with clinical trial approval in other countries.

However, IRBs have also in other countries been suggested as key stakeholder to influence the scientific fate of drug trials[40]. Thirdly, in particular regarding the reporting fate, publishers and editors are involved in the process of publishing trials.

With respect to the design of FIM trials, chapter 2.1 explained that pharmacology and dose-effect relations should be the core of their objectives, which is still often not the case. Sponsors and investigators can change this by using PK/PD approaches as much as possible in the determination of the first in man dose, choice of endpoints and escalating the dose[8,9]. Doing so is ethically imperative, as demonstrated by the trial tragedies of TGN1412 and BIA10-2474, where these approaches were wrongfully not used[41-43]. Furthermore, this may also change the modus operandi of drug development on a more fundamental level, as the structure of many drug development projects is currently focused on moving to phase 2 as soon as there is some preliminary indication of safety. Making phase 1 more data-rich requires more time and funding for this phase, but this investment can be earned back quickly by preventing futile phase 2 trials.

Regulators already stipulated the importance of PK/PD in guidelines[44,45], but they might be able to do more at the level of the IRB. IRBs can demand a structural justification of the available methods (the reason for using method X and not method Y and Z) underlying the design of a FIM trial. Doing so would fit within their mandate, as using the "right methodology" is formulated as the only condition under which human research is permitted in the law[46]. Thus, the proposed recommendations are already supported by prevailing legislation.

For each specific reason for (unjustified) discontinuation, a different solution might be available. Discontinuation due to recruitment failure can indicate (a combination of) several issues including insufficient availability or willingness of participants, resources, or recruitment time. Investigators should investigate such aspects before start of the trial by pilots and (financial) planning[47,48]. Furthermore, IRBs should critically scrutinize the justification of the sample size. If the justification indicates that the statistically required sample size is practically unfeasible, it can still be justified to approve the trial with a lower target sample size in a more exploratory design[48]. However, unfeasibility in obtaining and adequately big sample size requires in certain situations postponing IRB-approval until these issues have been solved. Next, discontinuation after an interim analysis demonstrated futility or superiority should always be preplanned in the trial protocol. Therefore, when designing trials involving formal statistical evaluations of efficacy and safety that involve interim testing, investigators should incorporate a plan in the trial protocol that describes the role of an independent Data Monitoring Committee, the timing of the interim analyses,

as well as the stopping rules[49,50]. This plan can then be evaluated by the IRB in the initial trial review. If the investigator or sponsor intends to discontinue the trial, independent and adequately planned review should demonstrate whether the intended discontinuation is justified.

The conventional way of reporting trial results is by publication of a peer-reviewed article in a scientific journal. Alternatives are uploading a summary of the results in a public trial registry or through presentation at scientific conferences. However, the latter option does usually not have sufficient time and space to disclose all relevant aspects and to use the results for designing future trials, evidence synthesis or clinical decision-making, leaving two options for results disclosure on the table. Investigators and sponsors should make sure that the complete picture of a clinical trial is disclosed through either means. This complete picture includes the original IRB-reviewed trial protocol and substantial amendments, so that the trajectory from the protocol until the publication becomes fully transparent. In the current era of online publishing, all journals offer options to attach web-only materials to publications, so space limitations should not be a problem.

IRBs can play an important role in achieving a full disclosure scenario, as they hold records of all clinical trials within their jurisdiction. With the coming EU-wide regulation[51] , a unique opportunity presents itself to ensure that the results of all trials are publically disclosed through the new digital infrastructure which is currently being developed for implementation of the new regulation. Publically disclosed information should at least contain the trial starting date, completion date of recruitment, completion date of follow-up, recruitment numbers, detailed data on design and methodology, all amendments, results, scientific and other publication. In this way, the totality of clinical research evidence, also about failures, will be available to inform clinical research and practice.

Public disclosure of all results of all trials is mandated by Dutch law. However, the law permits an escape, in case the sponsor of the trial formally objects to public disclosure. This escape is likely in place to protect intellectual property. This escape might be changed to an embargo time after which the results are still disclosed. If the drug product is effective, the company has then still sufficient time to market the drug under confidentiality of marketing-sensitive information, and the results will still be disclosed in case the development of the product is terminated. Furthermore, publishers and editors can also influence the disclosure of results, by stimulating and accepting manuscripts with statistically nonsignificant results and/or inconclusive findings. Recent research shows that journals accept trials with significant and non-significant results in equal rates, thus suggesting a positive trend[34].

A recommendation is to put more effort in the investigation and regulation of the relevance of the research question by conducting a systematic review before and after any RCT[52]. The review before the RCT needs to establish equipoise[7] and clinical relevance of the trial, and the review after the RCT examines how the results of the RCT have changed the overall body of knowledge. The presence and quality of such systematic reviews was not empirically investigated in this thesis, but others have convincingly demonstrated that it is too often not done[53-55]. If RCT protocols lack systematic reviews of existing literature, new RCTs may be redundant and also harmful and unethical, as the superiority of one of the trial arms was already established by previous RCTs. Allocation of patients to inferior treatments caused avoidable serious outcomes such as perioperative transfusions and pain[56,57]. Based on detailed systematic reviews, trial investigators can set a research agenda to inform further RCTs involving, for example, pediatric or geriatric populations, specific drug combinations or different dosage regimens.

These recommendations are not only intended for those working in the field of clinical trials, but particularly also for implementation in the training programs of new clinical researchers, drug developers, regulators, etcetera. It is still often being taught and written in pharmacology textbooks, for example, that safety and tolerability are the primary objectives of FIM trials[58]. So, regularly updating education materials according to the latest progress is needed. Furthermore, the importance of transparency, full disclosure and an unbiased literature needs emphasis throughout biomedical curricula. In this way, future generations of biomedical researchers will be properly trained to bring trials to a good fate.

Improving clinical trial governance and databases

Since its publication in 1996, the International Conference on Harmonization (ICH) GCP guideline and the subsequent national and supranational legislations based on it have received substantial amounts of criticism. Points of critique included that ICH-GCP has made the conduct of trials unnecessarily complicated, bureaucratic and expensive[59-61]; it did not harmonize regulations but rather made them more inconsistent[10,62]; on-site checking of source data is pointless and inefficient[63,64]; it obstructs the conduct of pragmatic trials[65]; and that it replaced quality by design for one-size-fits-all, box-ticking policies [66]. Indeed, quantitative studies demonstrate that clinical trials have become unnecessarily expensive and complicated[67,68]. Clinical researchers felt that they had been overlooked in devel-

---

[7]   A definition of equipoise was provided by Dr. Benjamin Freedman: "a state of genuine uncertainty on the part of the clinical investigator regarding the comparative therapeutic merits of each arm in a trial". (N Engl J Med 1987; 317: 141-5).

oping the guideline and that through the regulations, conducting clinical trials had become too much commercial business and too little science[69]. Fortunately, over the past years more dialogue took place, for example in the Clinical Trials Transformation Initiative (CTTI), focusing on quality by design approaches and more collaboration with the field.

The recommendation that follows from chapter 5 on risk-based supervisions and the points of critique as outlined above is that regulators and investigators should continue their dialogue on how to set up governance activities. Trials with a low risk (for example, testing products already available on the market) could be partially or completely waived from regulatory procedures. Furthermore, to minimize administrative costs, checking of data sources and informed consent leaflets should be reduced to a minimum, and done electronically as much as possible.

The Dutch law mandates, as in most countries, two main regulatory bodies responsible for supervision and oversight of clinical drug trials: the CCMO and the local accredited IRBs for the ethical review and the healthcare inspectorate to ensure compliance with the law. Furthermore, the Medicines Evaluation Board (MEB) is often involved in advising companies on which trials to conduct and which methodology to use to efficiently obtain marketing authorization. As it can be complicated for investigators and sponsors to comply with the plethora of regulatory requirements, regulations by the three agencies need careful calibration. Members of the licensing agency should be aware of the ethical review by the IRB, to prevent that they give advice that is ethically unfeasible.

To optimize the dialogue between regulators, access to each other's data may be a step in the right direction. This is already the case between the CCMO and the healthcare inspectorate, as the CCMO delivers periodical data reports to the inspectorate, which uses these reports to inform their inspection policy. There could also be more data sharing between CCMO and MEB, for example, the scientific advice reports and question letters following ethical review. In this way, both agencies are aware of communications between one agency and investigators or sponsors. If the CCMO identifies an issue in a scientific advice that can compromise the ethical review process, they can step in and discuss the issue, preventing a delayed IRB review process. For all stakeholders it is, in the end, most useful if government regulations are consistent and predictable.

While conducting the studies on the fate of clinical drug trials, we used regulatory and literature databases. Improving communication between these databases will likely be the most significant practical improvement in the governance of clinical trials. If these databases are improved, the total chain of regulatory agencies including national competent authorities, IRBs, healthcare inspector-

ates and drug marketing regulatory agencies and governmental policy makers and funders can efficiently monitor the performance and fate of clinical trials, so that the laborious research efforts as described in chapter 2 of this thesis are no longer needed. Based on the experiences gained during the studies, three main improvements could be advised for the regulatory database ToetsingOnline: 1) to improve functionality for data extractions; 2) to increase compliance with updating the status of trials, and 3) to introduce uniform indexing across the several clinical trial databases. These improvements should primarily aim to create a database which efficiently facilitates regulatory practice, as optimizing the scientific fate of clinical trials is part of the responsibilities of governance[70].

Regarding the first recommendation, easy data extractions will enable regulatory officers or regulatory researchers with the mandate to do so, to perform data extractions without the use of complicated programming or help by the IT expert. Regarding the second recommendation, updating of the status of trials means that all stages of progress of the trials are carefully monitored in the database. Although the status of trials can be entered in ToetsingOnline, it became clear during our studies that these fields were often missing, or were not updated. Moreover, some important updates should be added: in case of discontinuation, the reason for discontinuation, and the publication status, including reference to the location of the publication and the original trial protocol. Moreover, as the law requires it, governments should enforce reporting of results in public trial registries.

Regarding the third recommendation, regulatory and literature databases can be structured using uniform index terms. In ToetsingOnline, the identification number (ID) and index term is the NL-number. In EudraCT, this is the EudraCT-number. However, trials are usually not indexed by their IRB ID-numbers in widely used publication databases like Medline, Embase and Google Scholar, in the public registries clinicaltrials.gov and ISRCTN, and in the database of the MEB, that contains the regulatory product dossier. In particular, the trajectory of trials in the regulatory product dossier was difficult to reconstruct unambiguously by the IRB ID. This ID may actually be the most important index term, as it is assigned to a trial at its first regulatory exposure and all clinical trials go through IRB review. This will enable communication between the several academic and regulatory databases used for registration and reporting. As illustrated in figure 6.3, if this linkage can be established throughout the lifecycle of all trials, registrations, records, publications and other records of clinical trials can be unambiguously linked to the inception (IRB-review) of the trial. Such an ideal world scenario will incredibly improve the conduct of studies such as we performed in chapter 2, because all outcomes (fates) are readily available

through linkage of the several databases. Regulators and scientists could then collaborate on a global level to construct the complete picture of clinical evidence on a given drug, within a given period. The scenario depicted in figure 6.3 will not be easy to operationalize, in particular because trials are reviewed by IRBs across the globe[26]. In the globalized clinical drug research setting of today, a multicenter trial is reviewed by several IRBs around the world, which all have their own local database. Therefore, a concerted initiative is needed, to be supported by influential international organizations representing policy makers and scientists around the world.
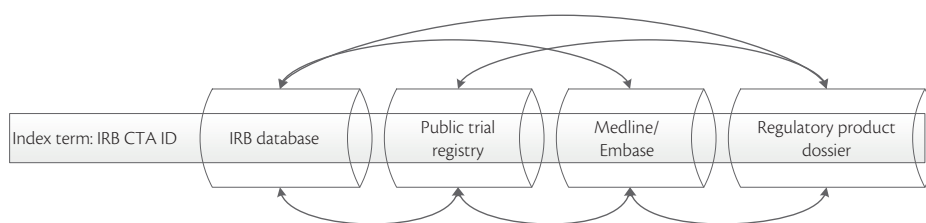

FIGURE 6.3 Ideal world scenario for indexing clinical trials: Institutional Review Board (IRB) clinical trial application (CTA) identification (ID) number is indexed throughout the whole lifecycle of the trial.

We also recommend some changes for the ABR-form. If improved, the ABR-form could become more useful for regulatory learning and risk-based supervision. First, the extensiveness of the public summary that is included in the ABR-form could be made less flexible. Now, the different parts are fixed (objectives, intervention, endpoints, etc.), but the investigators can still choose themselves how specific they describe, for example, the endpoints of the trial. Requiring the investigators to be as specific and complete as possible regarding the primary and secondary endpoints will improve the ABR-form as useful summary of the trial protocol. If this information becomes publically available, cases of selective outcome reporting can be readily identified, at least in an easier way compared to the assessment that we performed in chapter 2.4. Secondly, we recommend adding some trial-specific items to the ABR-form. Related to chapter 2.5, it would be useful if the ABR-form indicates whether the trial is intended for inclusion in a marketing authorization application. This gives regulators the opportunity to discuss with investigators whether the results of the trial are intended for submission to the regulatory product dossier. Another set of items could be added for the first in man (FIM) trials, as there are specific risks related to these trials. If the method of dose determination, intended dose cohorts and the type of data used to proceed to a higher dose are specified on the ABR-form, regulators could use these items as starting point for their assessment of the trial design and risks.

The proposed enhanced extraction functionality combined with no missing status updates and uniform indexing will enable all government and regulatory stakeholders to identify success and failure percentages, stratified for relevant characteristics such as trial site, sponsor, disease area, or trial phase. Inspectorates could easily identify sites where trials are often discontinued for unjustified reasons and hence GCP may be compromised, and funders can identify successful or promising new areas of pharmacotherapeutic development. Thus, improving the databases will provide detailed knowledge of successes and failures in clinical development, which are both important to inform government regulatory practice and strategies for public research funding. If they have databases that are more efficient at their disposal, governments may also be able streamline GCP-based regulations so that they will start to facilitate good clinical research.

For the EU, the new clinical trial regulation is an excellent opportunity to implement the recommended improvements, as a new trial application portal is currently being developed for its implementation. Although the new regulation only applies to clinical drug trials (approximately one-third of all clinical trials reviewed by Dutch IRBs), the new database portal could also be used for the non-drug trials, to keep the system as homogeneous as possible. Moreover, migrating the complete regulatory clinical trial database to an optimized database according to the latest standards will likely be a substantially larger one-time investment compared to incrementally improving ToetsingOnline, but will likely result in more value on the longer term.

PERSONAL REFLECTIONS

Much has been written about the importance of the fate of scientific research, that is, that all completed studies should be published in full[15,71,72]. This thesis contributes to this discussion by demonstrating that, indeed, waste due to poor completing and reporting rates is implicated in approximately 50% of the trials. This might seem paradoxical given the reward system in academia, where publications play an important role. Publishing an article in this system is usually an important milestone for researchers. However, curiously, the motivation not to publish a trial can apparently be stronger than the incentive of a scientific publication. Apparently, the disappointment and reluctance to write about a 'failed' trial can be strong to the extent that it outweighs the imperative to prevent others from making similar mistakes, to value the risks taken by trial participants, and to avoid distortion of the literature.

For commercial trials, reasons for not publishing negative results seem clear. Future investors are easily deterred by previous projects that did not deliver returns on investment. Furthermore, publications serve little purpose for pharmaceutical companies if this publication cannot be used for marketing their product. As phase 1 trials are mostly not useful in clinical practice, this probably explains why phase 1 trials are often not published, even if the results are positive. This leaves us with the question why the academic, investigator-initiated trials often do not reach the publication stage. Is it really the disappointment with the results that diminishes the motivation to write and submit an article? The findings of this thesis indicate that this is likely the largest, but not the only contributor. However, the predisposition of a trial for reaching an optimal or suboptimal fate is not only based on the direction of the results, but also on each of the stages from the framework of figure 1.1.

A possible underlying cause may be found within the academic system. The two most important reward metrics in the current system are citations and journal impact factors. The H-index, proposed by Professor Jorge Hirsch, is defined as "the number of papers with citation number ≥h, as a useful index to characterize the scientific output of a researcher"[73]. The journal impact factor is composed of a the number of citations in the current year to publications of the 2 previous years (the numerator) and the number of articles and reviews published in the same 2 years (the denominator)[74]. It is therefore important for the career of junior and mid-level scientist to attract citations, which can be achieved best by publishing in high-impact journals[75]. Publications of success-stories (in biomedical research, this often means statistically significant results) are known to attract more citations and be published in higher-impact factor journals than publications with nonsignificant and/or uncertain results. Before starting a trial, it is difficult to predict whether the trial will deliver the desired H-index and impact factor-increasing publication. Research groups may therefore rather take the risk of scientifically failure somewhere down the road and cut some corners during the preparations, such as establishing a good rationale, systematically summarizing all the available evidence and piloting. Publishing trials that are discontinued (likely resulting from the suboptimal preparations), or that have nonsignificant results, is lowered on the list of priorities, as they are probably ineligible for prestigious journals and hence the investigators focus on the next project to bring the desired high-impact publication. In the same way as one successful blockbuster drug is worthwhile many development failures, one blockbuster publication has more value for the career of a scientist than ten carefully written publications in moderate impact journals with nonsignificant results.

6

Another unintended contributor is the increasing growth of (biomedical) science as profession. Science has gradually evolved from an activity only available for the very top percentage of elite circles which could afford to spend time on research, via a community of a selected few intellectuals, to the current situation where access to higher education, or even free access[76], is viewed as a human right[77], with publically funded Bachelor's and Master's degree programs and substantial investments of public money in academic research[78]. That science is becoming an increasingly attractive career path is demonstrated by the 2.5-fold increase in PhD graduations in the Netherlands between 1990 and 2015[79]. This creates competition (availability of funding for only a very small percentage of applicants), and competition creates races with being first and fastest among their undesirable side effects. This competition is also present in the media, where journals struggle for readers, and journalists for being on the front-page. Deyo & Patrick[80], and also Mills[81], described how medical advances have been surrounded by unrealistic over-positivistic and hyped media attention. This is at clear odds with good science, which is a slow process requiring careful preparation, experimentation and replication. Under these current competitive circumstances, the business of science clearly conflicts with a responsible and transparent conduct of science. However, although investments in replication and full disclosure may not be as rewarding as it should be, it is the only way forward to improve, restore , or prevent loss of trust from the public in science[82] and solve its reproducibility crisis[83,84].

To increase the efficiency of the drug development process, adaptive pathways to marketing authorization have been proposed by the pharmaceutical industry and regulatory agencies[85,86]. Although giving patients earlier access to new drugs is desirable for severely ill patients[87], it should not lead to the idea that it is possible without randomization to generate similar evidence regarding the effectiveness of interventions[88]. Important in the implementation of adaptive pathways to marketing authorization is the use of 'real world evidence' and biomarkers in the approval process. Neither early phase biomarker trials nor observational studies are, however, suited to evaluate the effectiveness of a drug. Observational studies face unresolvable problems of alternative explanations in case of no effect, selection bias and unmeasured confounding[88,89]. Furthermore, although the role of biomarkers is important in many trial stages and aspects, they cannot replace clinical endpoints when it comes to informing clinical decision-making[90-93]. Using surrogate markers for effectiveness can result in harmful use of drugs, as shown by the antiarrhythmic drugs that were used against heart failure, but actually increased mortality in this population[94,95]. Moreover, the U.S. Food and Drug Administration published 22 recent cases where biomarkers did not predict

clinical endpoints[96]. In this light, RCTs evaluating clinical endpoints remain indispensable in establishing the effectiveness of new and existing drugs. The debate on regulatory reform should therefore not aim to find easier, cheaper or quicker ways to generate evidence similar or almost similar to the RCT, as this quest is doomed to fail.

Reforming the way in which drugs are regulated may nevertheless be inevitable. Restricting patients' access to drugs based on the presence of a government-approved data package is problematic, as this approach requires at least one of the assumptions that the benefit/risk profile of a drug can be translated to a uniform decision, taken on objective, scientific grounds, and that individual patients are incapable to make these decisions for themselves and should therefore be protected from making a 'wrong' decision. Both assumptions are unfounded. Individual patients clearly prioritize the risks and benefits of their daily life decisions in different ways, including drug use. In the translation of scientific evidence to a yes/no decision, is therefore ample room for personal and political factors[97]. Even if a majority of the population considers a given benefit/risk profile of a drug as negative, this is not a valid reason to withhold this drug from those with a different view. Moreover, there is no evidence that the average carpenter, truck driver or IT specialist is incapable to evaluate risks and benefits of therapies for their own specific situation. Clinical and epidemiological expertise of doctors and pharmacists is required to translate research findings to individual situations as good as possible. However, the decision should be in the hand of the patient, whether the government judges that the benefit/risk profile is positive or not. Thus, regardless how noble its intentions, the current regulatory system, yielding a residue of disapproved drug therapies and products, is undesirable. The government has an important, constitutional responsibility to provide this independent information as part of their task to promote public health[98]. Regulatory organizations such as the EMA have, therefore, a key role in judging the validity of marketing claims made in the label of drug products. Their scope should concern the dissemination of information, so that all individuals have access to and are able to identify this information. However, an end-decision of which therapies patients can and which they cannot access goes beyond these responsibilities. Thus, reform should be about who should be empowered to take risks regarding the treatments of individual patients and who should pay for them; a debate of the right to self-determination and providing patients with responsibility over their own decisions.

A similar line of argumentation may be used against the regulation of clinical trials. Before taking a drug, the user has (consciously or unconsciously) assessed the associated benefits and risks. This assessment of benefits (potentially effective

treatment, helping science, free complimentary healthcare, financial compensation) and risks (unknown side effects, additional hospital visits, burdensome medical procedures) is also performed by the trial participant before enrolling. However, to answer the research question, the study design and integrity of the collected data must be optimal. To avoid wasted or dangerous trials because of wrong design choices (including unjustified dose calculations), independent assessment by an expert committee such as an IRB of the scientific soundness remains required. For the assessment of benefits and risks associated with the trial, their proper role may be, similarly to marketing authorities, to demand complete and accurate provision of information to trial participants.

Notwithstanding the fact that RCTs are the only way to establish effectiveness of drugs, a problem is that many RCTs are not 'real-world-proof', jeopardizing their application in clinical practice[99]. A major improvement to most current RCTs, both pre- and post-marketing, would therefore be to increase the generalizability of the findings by increasing the representativeness of the studied population. The population on which drugs are tested, is too often a stringent selection from the population that is subsequently using the drugs. For example, patients included in conventional RCTs usually have shorter illness duration and higher tolerance against side effects[100,101]. The rationale of the concept of pragmatic trials, introduced in 1967 by Schwartz and Lellouch[99], is to conduct trials that resemble the actual clinical care situation in the actual target population that is using the drug. Still, limitations such as the willingness of patients and healthcare providers to participate in pragmatic trials[101-103], informed consent procedures[104,105], treatment blinding[106] and regulations[107-109] are barriers that prevent the conduct of pragmatic trials on a regular basis.

CONCLUSION

In modern drug research, the scientific fate of clinical trials is under pressure, as approximately half of them reach their optimal scientific fate. This optimal fate is a transparent report, describing all aspects of the trial needed to apply its findings in clinical practice, to use the findings in meta-analyses, and to inform other investigators in the design of future trials. Furthermore, the low success rates in clinical drug development call upon all stakeholders to establish a fostering environment for the conduct of clinical trials, as RCT are indispensable to test the efficacy and effectiveness of new and existing drugs. Regulatory clinical trial databases can be improved to monitor the scientific fate of trials on a meta-level in an ongoing and automated basis. The collaboration between the

different regulatory authorities involved in clinical drug trials and licensing can be improved by creating communicable databases and by adoption of risk-based approaches and quality by design. In the end, fully transparent, accountable and sound research is what it takes to truly advance medical innovation, stimulate participants to continue their collaboration with clinical research, and to enable medicine to deal with the challenges of this time.

EPILOGUE

Not long after the evens in the London hospital, Tegenero Immuno Therapeutics, the small German biotech company that was developing TGN1412, went bankrupt. An incident with an impact like this may be survived by a large pharmaceutical company, but not by a start-up with only R&D activities. Parexel, the multinational contract research organization that conducted the TGN1412 trial, settled with the volunteers over compensation and after the English regulatory authorities concluded that no then-current laws had been violated, Parexel could continue running trials for their sponsors.

The story did, however, not end then and there. In 2009, a new Russian biotech company named Theramab acquired the rights of TGN1412, and worked with the Institute for Virology and Immunobiology, led by professor Thomas Hünig and inventor of the antibody, on in vitro tests to elucidate the exact mechanism of action. TGN1412 was rebranded into TAB08 and in November 2011, a new clinical trial was successfully carried out in a Russian hospital[110]. The results indicated that TAB08, when given in the appropriate dose, could indeed selectively induce the regulatory immune system without inducing inflammation[111]. This success led to the initiation of 4 phase 2 clinical trials on patients with rheumatoid arthritis (clinicaltrials.gov identifier NCT01990157), psoriasis (NCT02796053), lupus erythematosus (NCT02711813), and solid tumors (NCT03006029), for which the results are being awaited. Regardless whether it will ever reach the clinic, the TGN1412/TAB08 story shows how the best design, proper conduct and full transparency in clinical research are essential to increase survival, not only for participants, but also for trials, companies, medicine and patients.

6

REFERENCES

1. Easterbrook PJ, Matthews DR. Fate of research studies. *J R Soc Med*. Feb 1992;85(2):71-76.
2. Pocock SJ. When to stop a clinical trial. *BMJ*. Jul 25 1992;305(6847):235-240.
3. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. Jul 04 2009;374(9683):86-89.
4. Paul SM, Mytelka DS, Dunwiddie CT, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov*. Mar 2010;9(3):203-214.
5. Otto MW, Nierenberg AA. Assay sensitivity, failed clinical trials, and the conduct of science. *Psychother Psychosom*. Sep-Oct 2002;71(5):241-243.
6. Marshall JC. Why have clinical trials in sepsis failed? *Trends Mol Med*. Apr 2014;20(4):195-203.
7. Harrison RK. Phase II and phase III failures: 2013-2015. *Nat Rev Drug Discov*. Dec 2016;15(12):817-818.
8. Cohen A. Should we tolerate tolerability as an objective in early drug development? *Br J Clin Pharmacol*. Sep 2007;64(3):249-252.
9. Cohen AF, Burggraaf J, van Gerven JM, Moerland M, Groeneveld GJ. The use of biomarkers in human pharmacology (Phase I) studies. *Annu Rev Pharmacol Toxicol*. 2015;55:55-74.
10. Al-Shahi Salman R, Beller E, Kagan J, et al. Increasing value and reducing waste in biomedical research regulation and management. *Lancet*. Jan 11 2014;383(9912):176-185.
11. Chalmers I, Bracken MB, Djulbegovic B, et al. How to increase value and reduce waste when research priorities are set. *Lancet*. Jan 11 2014;383(9912):156-165.
12. Chan AW, Song F, Vickers A, et al. Increasing value and reducing waste: addressing inaccessible research. *Lancet*. Jan 18 2014;383(9913):257-266.
13. Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. Jan 18 2014;383(9913):267-276.
14. Ioannidis JP, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*. Jan 11 2014;383(9912):166-175.
15. Dwan K, Altman DG, Clarke M, et al. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. *PLoS Med*. Jun 2014;11(6):e1001666.
16. Berendt L, Callreus T, Petersen LG, Bach KF, Poulsen HE, Dalhoff K. From protocol to published report: a study of consistency in the reporting of academic drug trials. *Trials*. 2016;17:100.
17. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*. May 26 2004;291(20):2457-2465.
18. Van Lent M, Rongen GA, Out HJ. Shortcomings of protocols of drug trials in relation to sponsorship as identified by Research Ethics Committees: analysis of comments raised during ethical review. *BMC Med Ethics*. Dec 10 2014;15:83.
19. Van Lent M. Influence of the funding source on quality of clinical drug trials. Nijmegen, the Netherlands: Radbout University; 2015.

20. Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA*. Aug 20 2003;290(7):921-928.

21. Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. *Cochrane Database Syst Rev*. Feb 16 2017;2:MR000033.

22. Eichberg MJ. Public funding of clinical-stage antibiotic development in the United States and European Union. *Health Secur*. May-Jun 2015;13(3):156-165.

23. Lewis-Fernandez R, Rotheram-Borus MJ, Betts VT, et al. Rethinking funding priorities in mental health research. *Br J Psychiatry*. Jun 2016;208(6):507-509.

24. Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat Rev Drug Discov*. Mar 01 2012;11(3):191-200.

25. Weber EJ, Callaham ML, Wears RL, Barton C, Young G. Unpublished research from a medical specialty meeting: why investigators fail to publish. *JAMA*. Jul 15 1998;280(3):257-259.

26. Glickman SW, McHutchison JG, Peterson ED, et al. Ethical and scientific implications of the globalization of clinical research. *N Engl J Med*. Feb 19 2009;360(8):816-823.

27. Rosenblatt M. The Large Pharmaceutical Company Perspective. *N Engl J Med*. Jan 05 2017;376(1):52-60.

28. Decullier E, Lheritier V, Chapuis F. Fate of biomedical research protocols and publication bias in France: retrospective cohort study. *BMJ*. Jul 2 2005;331(7507):19.

29. Kasenda B, von Elm E, You J, et al. Prevalence, characteristics, and publication of discontinued randomized trials. *JAMA*. Mar 12 2014;311(10):1045-1051.

30. Sune P, Sune JM, Montoro JB. Positive outcomes influence the rate and time to publication, but not the impact factor of publications of clinical trial results. *PLoS One*. 2013;8(1):e54583.

31. Arrowsmith J, Miller P. Trial watch: phase II and phase III attrition rates 2011-2012. *Nat Rev Drug Discov*. Aug 2013;12(8):569.

32. Berggren R, Moller M, Moss R, Poda P, Smietana K. Outlook for the next 5 years in drug innovation. *Nat Rev Drug Discov*. Jun 01 2012;11(6):435-436.

33. Reichert JM, Wenger JB. Development trends for new cancer therapeutics and vaccines. *Drug Discov Today*. Jan 2008;13(1-2):30-37.

34. Evoniuk G, Mansi B, DeCastro B, Sykes J. Impact of study outcome on submission and acceptance metrics for peer reviewed medical journals: six year retrospective review of all completed GlaxoSmithKline human drug research studies. *BMJ*. Apr 21 2017;357:j1726.

35. Odutayo A, Copsey B, Dutton S, Cook J, Hopewell S, Altman DG. Characteristics and Dissemination of Phase 1 Trials Approved by a UK Regional Office in 2012. *JAMA*. May 02 2017;317(17):1799-1801.

36. Piller C. Law ignored, patients at risk. Accessed 15 May 2017. . http://www.statnews.com/2015/12/13/clinical-trials-investigation/: Statnews.com; 2015.

37. Hume D. *A Treatise of Human Nature; Part 3: Of the will and direct passions*. London, UK1738.

38. Peeters PJHL. *Etiological studies in complex diseases*. Utrecht, the Netherlands: Utrecht Institute for Pharmaceutical Sciences, Utrecht University; 2016.

39. Goldacre BD, H.; Powell-Smith, A. The COMPare Trials Project (http://compare-trials.org/). Oxford, UK: Nuffield Department of Primary Care Health Sciences; 2016.

6

40. Doernberg SN, Wendler D. Ensuring Respect for Human Research Participants: Institutional Review Boards and Sharing Results From Research. *JAMA.* Sep 20 2016;316(11):1149-1150.

41. Expert Scientific Group mandated by the Secretary of State for Health. Phase One Clinical Trials Final Report. London, UK: The Stationery Office; 2006.

42. Kenter MJ, Cohen AF. Establishing risk of human experimentation with drugs: lessons from TGN1412. *Lancet.* Oct 14 2006;368(9544):1387-1391.

43. The Temporary Specialist Scientific Committee "FAAH (Fatty Acid Amide Hydrolase)". Report on the causes of the accident during a Phase 1 clinical trial in Rennes in January 2016. . Saint Denis, France: Agence Nationale de Sécurité du Médicament et des produits de santé; 2016.

44. European Medicines Agency. Concept paper on the revision of the Guideline on strategies to identify and mitigate risks for first-in-human clinical trials with investigational medicinal products (EMA/CHMP/446302/2016). London, UK: EMA; 2016.

45. Food and Drug Administration. Guidance for Industry: Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers. Silver Spring, Maryland USA: U.S. Food and Drug Administration; 2005.

46. Minister of Health Welfare and Sports of the Netherlands. Medical Research Involving Human Subjects Act (in Dutch: Wet medisch-wetenschappelijk onderzoek met mensen). The Hague, the Netherlands: Staatsblad van het Koninkrijk der Nederlanden; 2017.

47. Briel M, Olu KK, von Elm E, et al. A systematic review of discontinued trials suggested that most reasons for recruitment failure were preventable. *J Clin Epidemiol.* Aug 3 2016.

48. Carlisle B, Kimmelman J, Ramsay T, MacKinnon N. Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials. *Clin Trials.* Feb 2015;12(1):77-83.

49. Committee for Medicinal Products for Human Use (CHMP). Guideline on Data Monitoring Committees. EMEA/CHMP/EWP/5872/03. London, UK: EMA; 2006.

50. International Conference on Harmonisation E9 Expert Working Group. ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. Geneva, Switzerland: ICH; 1999.

51. European Commission. Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC.: European Commission; 2014.

52. Clarke M, Hopewell S, Chalmers I. Clinical trials should begin and end with systematic reviews of relevant evidence: 12 years and waiting. *Lancet.* Jul 03 2010;376(9734):20-21.

53. Clarke M, Brice A, Chalmers I. Accumulating research: a systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. *PLoS One.* 2014;9(7):e102670.

54. Jones AP, Conroy E, Williamson PR, Clarke M, Gamble C. The use of systematic reviews in the planning, design and conduct of randomised trials: a retrospective cohort of NIHR HTA funded trials. *BMC Med Res Methodol.* Mar 25 2013;13:50.

55. Sheth U, Simunovic N, Tornetta P, 3rd, Einhorn TA, Bhandari M. Poor citation of prior evidence in hip fracture trials. *J Bone Joint Surg Am.* Nov 16 2011;93(22):2079-2086.

56. Fergusson D, Glass KC, Hutton B, Shapiro S. Randomized controlled trials of aprotinin in cardiac surgery: could clinical equipoise have stopped the bleeding? *Clin Trials.* 2005;2(3):218-229; discussion 229-232.

57. Habre C, Tramer MR, Popping DM, Elia N. Ability of a meta-analysis to prevent redundant research: systematic review of studies on pain from propofol injection. *BMJ.* Aug 26 2014;348:g5219.

58. Lundqvist T, Bredenberg S. *Pharmaceutical development (chapter 17, p. 242-244) in: Drug Discovery & Development - Technology in transition.* Second edition ed. London, UK: Elsevier Ltd.; 2012.

59. Butler J, Fonarow GC, O'Connor C, et al. Improving cardiovascular clinical trials conduct in the United States: recommendation from clinicians, researchers, sponsors, and regulators. *Am Heart J.* Mar 2015;169(3):305-314.

60. McMahon AD, Conway DI, Macdonald TM, McInnes GT. The unintended consequences of clinical trials regulations. *PLoS Med.* Nov 2009;3(11):e1000131.

61. Yusuf S. Randomized clinical trials: slow death by a thousand unnecessary policies? *Cmaj.* Oct 12 2004;171(8):889-892; discussion 892-883.

62. Hartmann M. Impact assessment of the European Clinical Trials Directive: a longitudinal, prospective, observational study analyzing patterns and trends in clinical drug trial applications submitted since 2001 to regulatory agencies in six EU countries. *Trials.* Apr 29 2012;13:53.

63. Baigent C, Harrell FE, Buyse M, Emberson JR, Altman DG. Ensuring trial validity by data quality assurance and diversification of monitoring methods. *Clin Trials.* 2008;5(1):49-55.

64. Bakobaki JM, Rauchenberger M, Joffe N, McCormack S, Stenning S, Meredith S. The potential for central monitoring techniques to replace on-site monitoring: findings from an international multi-centre clinical trial. *Clin Trials.* Apr 2012;9(2):257-264.

65. Mentz RJ, Hernandez AF, Berdan LG, et al. Good Clinical Practice Guidance and Pragmatic Clinical Trials: Balancing the Best of Both Worlds. *Circulation.* Mar 01 2016;133(9):872-880.

66. MoreTrials Campaign. ICH-GCP guideline: What are the major problems with the ICH-GCP guideline? http://moretrials.net/the-problem/gcp/. Accessed 15 May 2017.

67. Getz KA, Wenger J, Campo RA, Seguine ES, Kaitin KI. Assessing the impact of protocol design changes on clinical trial performance. *Am J Ther.* Sep-Oct 2008;15(5):450-457.

68. Hind D, Reeves BC, Bathers S, et al. Comparative costs and activity from a sample of UK clinical trials units. *Trials.* May 02 2017;18(1):203.

69. Campaign M. Updated open Letter to EMA & ICH: From 5 research organisations and an international consortium of 119 health researchers in 22 countries. Accessed 16 May 2017. http://moretrials.net/the-problem/2016.

70. Zarin DA, Tse T, Williams RJ, Carr S. Trial Reporting in ClinicalTrials.gov - The Final Rule. *N Engl J Med.* Nov 17 2016;375(20):1998-2004.

71. Glasziou P, Chalmers I. Is 85% of health research really "wasted"? Accessed 6 June 2017. http://blogs.bmj.com/bmj/2016/01/14/paul-glasziou-and-iain-chalmers-is-85-of-health-research-really-wasted/: The BMJ Opinion; 2016.

72. Glasziou P, Chalmers I. Can it really be true that 50% of research is unpublished? Accessed 6 June 2017. http://blogs.bmj.com/bmj/2017/06/05/paul-glasziou-and-iain-

6

chalmers-can-it-really-be-true-that-50-of-research-is-unpublished/: The BMJ Opinion; 2017.

73. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A.* Nov 15 2005;102(46):16569-16572.

74. Garfield E. The history and meaning of the journal impact factor. *JAMA.* Jan 04 2006;295(1):90-93.

75. Bornmann L, Daniel H-D. Does the h-index for ranking of scientists really work? *Scientometrics.* 2005;65(3):391-392.

76. General Assembly of the United Nations. Resolution 2200A (XXI): International Covenant on Economic, Social and Cultural Rights. New York, USA: United Nations; 1966.

77. General Assembly of the United Nations. Resolution 217A: The Universal Declaration of Human Rights. Paris, France: United Nations; 1948.

78. Rathenau Instituut. Overheidsfinanciering van R&D (Dutch). https://www.rathenau.nl/nl/page/overheidsfinanciering-van-rd: Rathenau Instituut. Accessed 2 May 2017; 2016.

79. Onderwijs in Cijfers (Education in Figures). Ontwikkeling van het aantal gepromoveerden (Dutch only). https://www.onderwijsincijfers.nl/kengetallen: Onderwijs in Cijfers. Accessed 1 May 2017; 2017.

80. Deyo RA, Patrick DL. *Hope or hype : the obsession with medical advances and the high cost of false promises. Chapter 9: "Cancer Cured - Film at 11:00".* New York: AMACOM, American Management Association; 2005.

81. Mills JL. Reporting provocative results. Can we publish 'hot' papers without getting burned? *JAMA.* Dec 18 1987;258(23):3428-3429.

82. Getz KA. Getting to the Heart of Public Distrust. http://www.appliedclinicaltrialsonline.com/getting-heart-public-distrust: Applied Clinical Trials; 2008.

83. Bouter LM. Open data are not enough to realize full transparency. *J Clin Epidemiol.* Feb 2016;70:256-257.

84. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature.* May 25 2016;533(7604):452-454.

85. Eichler HG, Baird LG, Barker R, et al. From adaptive licensing to adaptive pathways: delivering a flexible life-span approach to bring new drugs to patients. *Clin Pharmacol Ther.* Mar 2015;97(3):234-246.

86. European Medicines Agency. Final report on the adaptive pathways pilot (EMA/276376/2016). London, UK: EMA; 2016.

87. Baird LG, Banken R, Eichler HG, et al. Accelerated access to innovative medicines for patients in need. *Clin Pharmacol Ther.* Nov 2014;96(5):559-571.

88. Davis C, Lexchin J, Jefferson T, Gotzsche P, McKee M. "Adaptive pathways" to drug authorisation: adapting to industry? *BMJ.* Aug 16 2016;354:i4437.

89. Bosco JL, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol.* Jan 2010;63(1):64-74.

90. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. *Stat Med.* Nov 10 2012;31(25):2973-2984.

91. Chokshi DA, Avorn J, Kesselheim AS. Designing comparative effectiveness research on prescription drugs: lessons from the clinical trial literature. *Health Aff (Millwood).* Oct 2010;29(10):1842-1848.

92. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ.* May 08 2015;350:h2147.

93. Murad MH, Shah ND, Van Houten HK, et al. Individuals with diabetes preferred that future trials use patient-important outcomes and provide pragmatic inferences. *J Clin Epidemiol.* Jul 2011;64(7):743-748.

94. The Cardiac Arrhythmia Suppression Trial II Investigators. Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. *N Engl J Med.* Jul 23 1992;327(4):227-233.

95. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med.* Mar 21 1991;324(12):781-788.

96. Food and Drug Administration. 22 Case Studies Where Phase 2 and Phase 3 Trials had Divergent Results. Silver Spring MD, USA: FDA; 2017.

97. May RJ. Swamped By Science? http://www.pff.org/issues-pubs/other/opinion/040119legaltimes.html. Legal times; 2004. Accessed 3 June 2017.

98. Constitution for the Kingdom of the Netherlands, revision 15 July 2008.

99. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis.* Aug 1967;20(8):637-648.

100. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials.* Nov 03 2015;16:495.

101. Oude Rengerink K, Kalkman S, Collier S, et al. Participant eligibility, recruitment and retention in pragmatic trials. *J Clin Epidemiol.* May 11 2017.

102. Worsley SD, Oude Rengerink K, Irving E, et al. Challenges in Pragmatic Trials: Selection and Inclusion of Usual Care Sites. *J Clin Epidemiol.* May 11 2017.

103. Van Staa TP, Dyson L, McCann G, et al. The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials. *Health Technol Assess.* Jul 2014;18(43):1-146.

104. Kalkman S, van Thiel GJ, Zuidgeest MG, et al. Challenges of informed consent for pragmatic trials. *J Clin Epidemiol.* May 11 2017.

105. McKinney RE, Jr., Beskow LM, Ford DE, et al. Use of altered informed consent in pragmatic clinical research. *Clin Trials.* Oct 2015;12(5):494-502.

106. Welsing P, Oude Rengerink K, Collier S, et al. Outcome Selection and Measurement in Pragmatic Trials. *J Clin Epidemiol.* May 11 2017.

107. Califf RM, Sugarman J. Exploring the ethical and regulatory issues in pragmatic clinical trials. *Clin Trials.* Oct 2015;12(5):436-441.

108. Sugarman J, Califf RM. Ethics and regulatory complexities for pragmatic clinical trials. *JAMA.* Jun 18 2014;311(23):2381-2382.

109. Irving E, van den Bor R, Welsing P, et al. Collecting and Reporting Safety Data and Monitoring Trial Conduct In Pragmatic Trials. *J Clin Epidemiol.* May 11 2017.

110. Kenter MJ, Cohen AF. The return of the prodigal son and the extraordinary development route of antibody TGN1412 - lessons for drug development and clinical pharmacology. *Br J Clin Pharmacol.* Apr 2015;79(4):545-547.

111. Tabares P, Berr S, Romer PS, et al. Human regulatory T cells are selectively activated by low-dose application of the CD28 superagonist TGN1412/TAB08. *Eur J Immunol.* Apr 2014;44(4):1225-1236.

6

Chapter 7

Summaries

Chapter 7.1

English summary

Chapter 1 describes how in modern medicine, (randomized) clinical trials have evolved to essential instruments to establish evidence on interventions, such as medicinal products (drugs). The scientific fate of each clinical trial should be that a relevant research question is answered. This can be done by application of appropriate design, conduct and reporting in the academic literature describing all pre-planned and post-hoc results. In addition to the scientific fate, clinical drug trials can also have a development fate. The development fate concerns whether the knowledge gained by the trial is successfully incorporated in licensed, marketable product claims. The scientific fate should be independent from the direction or magnitude of the results of the trial, whereas the development fate is determined by the results of the trial. The first research question of this thesis concerns the scientific and development fate of clinical drug trials in the Netherlands.

Governments have implemented legislation following the Good Clinical Practice (GCP) guideline (1996), to safeguard the safety of trial participants and the integrity and reliability of trial data. Policies of the agencies that are mandated with regulation of GCP legislation (i.e. institutional review boards, inspectorates) can use risk-based approaches to concentrate their workforce on the trials with an elevated risk for issues in safety or data integrity, and to reduce bureaucracy for less risky trials. Routinely collected data by regulatory agencies may facilitate such risk-based approaches. The second research question of this thesis focuses on indicators that can facilitate risk-based clinical trial governance, and the role of regulatory databases in this process.

In chapter 2, a series of analyses is described of a cohort comprising clinical drug trials reviewed in 2007 by all Institutional Review Boards (IRBs) within the Netherlands, aiming to identify the scientific and development fate of these trials. Chapter 2 is divided into five sub-chapters, dealing with the design (2.1), completion (2.2), reporting (2.3 and 2.4) and licensing lifecycle (2.5) of the included trials.

In chapter 2.1, the role of pharmacology in the design of first in man (FIM) trials was investigated. The results showed that pharmacology was not considered in a substantial proportion of FIM protocols. Thirty-eight percent of the FIM trials used only the no observed adverse effect level as preclinical information for the dose calculation, and 31% applied allometric scales without discussion of potential interspecies differences. The majority of FIM trials (84%) measured, in addition to safety, tolerability and pharmacokinetics, also pharmacodynamic parameters. Thirty-eight percent of the FIM trials with a dose escalation design based the dose escalation only on safety/tolerability endpoints. Only a limited improvement was observed between FIM trials approved in 2007 and FIM tri-

7

als approved in 2015. Based on these findings, it was concluded that FIM trial protocols often lack thorough discussion of important pharmacological aspects in the calculation of the first dose and in their dose escalation procedures. If these aspects are ignored, dose-related problems could occur such as serious incidents during the FIM trial or overdose issues later in the lifecycle of the drug.

In chapter 2.2, the occurrence of and reasons for discontinuation of clinical drug trials before the planned recruitment or end of follow-up was examined, demonstrating that 102 out of the 574 (18%) analyzed trials were discontinued. The most common reason was recruitment failure (33/574, 5.7%), followed by interim analysis demonstrating futility of one of the trial arms (30/574, 5.2%). Investigator-initiated trials were more often discontinued due to recruitment failure compared to industry-sponsored trials. With respect to discontinuation after interim analysis, it was found that 10 (33.3%) of the trials contained an adequate interim analysis plan in the trial protocol. In total, 69 (12%) trials of the cohort were discontinued due to unjustified and/or undesirable reasons. The conclusion is, therefore, that trials discontinue too often due to questionable reasons, and that a substantial part of these discontinuations could be avoided by a better feasibility assessment of the sample size and by a better planning of interim analyses in trial protocols.

Chapter 2.3 describes a study that aimed to assess the proportion of clinical drug trials in the cohort that were published as peer-reviewed journal article. Out of the 574 trials analyzed, 334 (58%) trials were published after 8-9 years of IRB-review. Phase 1 (except oncology), not prospectively registered, single center and discontinued trials were statistically significantly associated with non-publication. Following-up on this study, chapter 2.4 zoomed in on the published trials by comparing the original IRB-approved trial protocols with the identified publications. The purpose of this comparison was to identify undisclosed protocol-publication discrepancies in primary and secondary endpoints, primary and secondary objectives, inclusion and exclusion criteria, sample size, sponsorship acknowledgement, subgroup analyses, and methods used for data analysis. Discrepancies were found in primary endpoints among 10%, in secondary endpoints among 41%, in primary objectives among 8%, in secondary objectives among 19%, in inclusion and exclusion criteria among 11%, in sample size among 11%, in subgroup analysis among 45%, and in methods used for data analysis among 10% of the cohort. Investigator-initiated trials, non-randomized and/or exploratory trials, not prospectively registered trials, and trials other than phase 1, 2, 3, or 4 were associated with having protocol-publication discrepancies in the primary endpoints. The conclusion of chapters 2.3 and 2.4 is that clinical trial results are still often not or selectively published in the literature, and

consequently that more transparency is required to reduce research waste and publication bias that is implicated with non-publication of study results.

The series of analyses of the 2007 cohort is concluded by chapter 2.5, describing an analysis of submission of clinical trial results from the 2007 cohort to the product dossier of the drug licensing authorities (European Medicines Agency (EMA) for Europe-wide licensing applications, Medicines Evaluation Board for decentralized license applications). It was found that 197 out of 558 (35%) eligible trials were submitted within 9-10 years after IRB-review. Reasons that trials were not submitted included that the product was not licensed (yet) for the EU or Dutch market (183/558, 33%); because the trial was not sponsored by the manufacturer of the product of interest (153/558, 27%); because the investigated indication or dose was not licensed (yet) for the EU or Dutch market (14, 3%); or because the trial results were in line with already registered label claims (11/558, 2%). The conclusion of this chapter is that no legally required trials were missing in the product dossiers, but that investigator-initiated trials that are not being included in the current system might be a potentially valuable data source for the regulatory product dossier and licensed label. Pathways independent from the manufacturers may therefore be needed to involve these trials in drug licensing.

In chapter 3, a systematic review is described of the scientific literature on determinants of selective reporting. Using quantitative and qualitative content analysis, a taxonomy was developed of the different actors and factors that contribute to selective reporting. The categories of this taxonomy were used to derive a hypothetical causal scheme elucidating the mechanisms behind selective reporting. Twelve determinant categories were established. In the causal scheme, these 12 categories were linked to two necessary causes including motives and means (at least one of these need to be at hand to cause selection bias) and two component causes including conflicts of interests and environmental pressures (these are not sufficient in themselves but contribute to the bias that leads to selective reporting). The taxonomy and causal scheme could help the development of and research on interventions to stimulate responsible research practices and reduce selective reporting.

Chapter 4 focuses on the global context in which clinical drug trials are conducted. In chapter 4.1, a comparison is described between the total phase 3 clinical trial activity within the European Union in the fields of oncology (the field with many pharmacotherapeutic successes) and psychiatry (a field with very few pharmaceutical development success over the past years). The results show that the magnitude of phase 3 clinical trial activity is related to the pharmacotherapeutic success in that area. In addition, the importance of maintaining

7

phase 3 clinical trial activity within the European Union and improvements to the quality of phase 3 trials independent of the disease area are discussed.

In chapter 4.2, the relationship is studied between partnering during clinical development and success of the marketing authorization application (MAA) outcome of new active substances (NAS). A cohort of all NAS MAAs between 2009 and 2013 of NAS were analyzed whether any mergers, acquisitions or in-licensing occurred during clinical development. Out of the 172 NAS, 133 (77%) were approved by the EMA. The percentage of approvals was higher among medium/large-sized company applicants (108/130, 83%) compared to small-sized company applicants (25/42, 60%). Furthermore, NAS originated by small companies were more likely to be approved if the company or NAS was acquired during clinical development (37/50 approved, 74%) than if no acquisition or other partnering occurred (17/27 approved, 63%). These findings led to the conclusion that for small companies, partnering with medium- or large-sized pharmaceutical companies is a good strategy to bring new products successfully to the market.

Chapter 5 primarily addresses the research question of risk-based approaches to GCP governance. Starting points were a risk model developed for the healthcare inspectorate and the input of a brainstorm session with participation of senior regulators from the Dutch regulatory authorities. Based on this risk model and brainstorm session, a further review of the literature and regulatory guidelines was conducted to identify readily recognizable risk indicators that indicate that the risk for issues in safety and/or data integrity is potentially increased. The identified risk indicators were incorporated in a taxonomy with a tree-structure. The three first risk branches included 'what' (indicators concerning risks associated with the products tested), 'by whom' (indicators concerning risks associated with the responsible parties conducting the trial), and 'how' (indicators concerning risks associated with design and methodological aspects of the trial). The risk indicator taxonomy can be used to facilitate a structured identification and analysis of risks for GCP issues. Stakeholders should prioritize risks depending on their perspective and task. Furthermore, the indicators are meant to be a valuable tool to supplement (and not to replace) experience- and context-driven intuition of expert trial supervisors.

As the final chapter, the general discussion (chapter 6) ties the findings on the fate of the cohort of drug trials together. It describes how of the 622 IRB-reviewed trials in the Netherlands in 2007, 283 (45%) reach their optimal scientific fate, and discusses the different degrees of the other scientific fates that can be considered as 'less than optimal'. Furthermore, chapter 6 briefly touches upon the limitations of the used methodology (i.e. no causal inference can be made from observational research), and provides a framework for the interpretation

of the investigated determinants. The main suggestion for further research is to identify the whole lifecycle of a cohort of products with regard to success and failure in drug development. Most current research takes a rather fragmented approach, such as looking at marketing authorization applications or pharmacovigilance signals (regulatory perspective), sales (business perspective), or utilization and adherence (clinical perspective). Combining these aspects could lead to a richer learning process about the drug products that enter the market. The main suggestions for practice concern what the different stakeholders can do to improve the design, conduct and reporting of clinical trials. Also, specific recommendations are provided for the efficient use of regulatory databases that enable a straightforward identification of the fate of clinical trials.

7

Chapter 7.2

Dutch summary
(Nederlandse samenvatting)

Hoofdstuk 1 legt uit dat klinische studies in de moderne geneeskunde van groot belang zijn in de wetenschappelijke vaststelling van de effectiviteit van interventies, waaronder geneesmiddelen. De gewenste wetenschappelijke bestemming van elke studie is een antwoord op de onderzoeksvraag door middel van een goede studieopzet, uitvoering en rapportage van alle resultaten in de wetenschappelijke literatuur. Naast de wetenschappelijke bestemming heeft een geneesmiddelenstudie vaak ook een beoogde ontwikkelingsgerichte bestemming, namelijk dat de kennis die door de studie wordt verworven, wordt opgenomen in de geregistreerde productinformatie. De wetenschappelijke bestemming van een klinische studie is onafhankelijk van de richting en omvang van de studieresultaten, terwijl de ontwikkelingsgerichte bestemming juist afhankelijk is van de resultaten. De eerste onderzoeksvraag die in dit proefschrift wordt behandeld gaat over welk deel van de klinische geneesmiddelenstudies uiteindelijk een optimale wetenschappelijke en ontwikkelingsgerichte bestemming heeft.

Ter bescherming van het welzijn van de proefpersonen en de integriteit van onderzoeksgegevens hebben nationale en supranationale overheden wetgeving ingesteld, gestoeld op het in 1996 gepubliceerde richtsnoer Goede Klinische Praktijken. De gemandateerde overheidsinstanties kunnen een risicogestuurde aanpak hanteren in de regulering en toezicht op de naleving van deze wetgeving. Het beoogde voordeel van een risicogestuurde aanpak is dat meer aandacht kan uitgaan naar de daadwerkelijk risicovolle studies, terwijl studies met een lager risico juist minder belast worden met wet- en regelgeving. Routinematig verzamelde gegevens door overheidsinstanties betrokken bij de regulering en toezicht van klinische studies zijn mogelijk waardevol in de implementatie van een risicogestuurde aanpak. De tweede onderzoeksvraag van dit proefschrift richt zich daarom op het identificeren van indicatoren voor risicogestuurde aanpak van regulering en toezicht, en hoe de routinematig verzamelde gegevens van de uitvoerende overheidsinstanties hiervoor kunnen worden gebruikt.

Hoofstuk 2 bevat een serie studies naar de wetenschappelijke en ontwikkelingsgerichte bestemmingen van klinische geneesmiddelenstudies aan de hand van een cohort van alle klinische geneesmiddelenstudies die in 2007 door de Nederlandse erkende medisch-ethische toetsingscommissies (METC's) zijn goedgekeurd. Het hoofdstuk is opgedeeld in 5 sub-hoofdstukken die de opzet, afronding, rapportage en registratie van de geïncludeerde trials behandelen.

In hoofdstuk 2.1 werd de opzet bestudeerd van klinische geneesmiddelenstudies die een nieuw geneesmiddel voor het eerst in mensen testten. Uit de resultaten blijkt dat een substantieel deel van de protocollen van deze studies relevante farmacologische overwegingen achterwege laat. Achtendertig procent van de studies baseerde de eerste humane dosering alleen op de 'no observed adverse

effect level' vanuit de preklinische tests, en 31% berekende deze dosering aan de hand van allometrische schalen zonder de aan- of afwezigheid van relevante verschillen tussen mens en dier te bespreken. Naast de veiligheid/verdraagbaarheid en farmacokinetiek werden farmacodynamische parameters in 84% van de studies gemeten. Echter, 38% van de studies die dosisescalatie toepasten baseerde dosisverhoging uitsluitend op veiligheid en verdraagbaarheid (en dus niet op farmacokinetische of farmacodynamische parameters). De vergelijking tussen de studies uit 2007 en 2015 liet slechts een beperkte verbetering zien. Op basis van deze bevindingen is de conclusie van dit hoofdstuk dat studieprotocollen van klinische studies waarin een geneesmiddel voor het eerst in mensen wordt getest, vaak onder de maat zijn wat betreft overwegingen van essentiële farmacologische aspecten. Het negeren van deze aspecten verhoogt het risico op zowel ernstige incidenten tijdens de klinische studie als op overdosering tijdens de verdere levenscyclus van het geneesmiddel.

In hoofdstuk 2.2 werd onderzocht welke studies in het cohort de geplande inclusie en looptijd volledig doorlopen hebben en welke voortijdig zijn beëindigd. Het bleek dat 102 van de 574 studies (18%) voortijdig waren beëindigd. Het falen van inclusie was de vaakst gerapporteerde reden voor voortijdige beëindiging (33 van de 574, 6%), gevolgd door het voortijdig aantonen van inferioriteit van één van de behandelarmen door een interim-analyse (30 van de 574, 5%). Onderzoeker-geïnitieerde studies hadden een hogere kans op voortijdige beëindiging vanwege falende inclusie. Met betrekking tot de interim-analyses had 10 van de 30 (33%) studies deze adequaat beschreven in het studieprotocol. In totaal zijn 69 van de 574 (12%) studies voortijdig beëindigd vanwege een reden die als ongefundeerd of twijfelachtig beschouwd kan worden. De conclusie van dit subhoofdstuk is dat studies te vaak voortijdig eindigen vanwege twijfelachtige redenen. Een aanzienlijk deel van dergelijke beëindigingen kan in de toekomst waarschijnlijk worden voorkomen door het beter schatten van de haalbaarheid van de steekproefgrootte en een betere planning van de interim-analyse.

Hoofdstuk 2.3 beschrijft hoeveel studies uit het cohort zijn gepubliceerd in de wetenschappelijke literatuur, 8-9 jaar na goedkeuring door de METC, en wat determinanten zijn van non-publicatie. Van de 574 studies in de analyse waren er 334 (58%) gepubliceerd. Een verhoogd risico op non-publicatie hadden fase 1-, monocentrische, niet prospectief geregistreerde en voortijdig beëindigde studies. In hoofdstuk 2.4 werden de 334 gepubliceerde studies verder onder de loep genomen. De onderzoeksprotocollen werden vergeleken met de geïdentificeerde publicaties op discrepanties in belangrijke rapportage-aspecten. Van de 334 bleken 32 studies (10%) een onverklaarbare discrepantie te hebben tussen de geprotocolleerde primaire eindpunten en de primaire eindpunten zoals beschreven in de

publicatie(s). Daarnaast hadden 41% van de studies discrepanties in secundaire eindpunten, 8% in de primaire onderzoeksdoelen, 19% in secundaire onderzoeksdoelen, 11% in selectiecriteria voor de onderzoekspopulatie, 11% in de grootte van de steekproef, 45% in subgroep-analyses, en 10% in de gehanteerde methode voor data-analyse. Op basis van hoofdstuk 2.3 en 2.4 kan worden geconcludeerd dat het selectief rapporteren van studieresultaten nog altijd vaak voorkomt, en dat daarom meer transparantie vereist is om verspilling van onderzoeksgelden en vertekening in de literatuur te voorkomen.

De onderzoeksvraag van hoofdstuk 2.5, de laatste studie met het 2007-cohort, was hoeveel studies uit het cohort uiteindelijk worden geïncludeerd in een productdossier van het Europees Geneesmiddelenagentschap (EMA; voor Europese registraties) of van het College ter Beoordeling van Geneesmiddelen (voor gedecentraliseerde registraties). Er bleken 197 van de 558 in aanmerking komende geneesmiddelenstudies (35%) te zijn geïncludeerd in het productdossier. De overige studies waren niet geïncludeerd omdat er (nog) geen registratiedossier bestond voor het product (33%), omdat de studie niet was gesponsord door de fabrikant van het product (27%), omdat de dosering of indicatie niet was geregistreerd (3%), of omdat de bevindingen van de studie in lijn waren met reeds geregistreerde claims (2%). Hoofdstuk 2.5 concludeert dat er geen aanwijzingen waren dat wettelijk verplichte studiedata niet in de dossiers aanwezig was. Echter, onderzoeker-geïnitieerde studies kunnen belangrijke informatie bevatten met betrekking tot de geregistreerde productclaims en zouden daarom vaker als databron gebruikt kunnen worden voor het productdossier. Onafhankelijk van indiening door de fabrikanten zouden deze studies in de toekomst betrokken kunnen worden bij het registratieproces.

Hoofdstuk 3 beschrijft een systematische overzichtsstudie van de literatuur over de oorzaken van selectief rapporteren. Door kwantitatieve en kwalitatieve analyse van de inhoud van de gevonden artikelen werd een taxonomie opgesteld van de verschillende actoren en factoren die een rol spelen in selectief rapporteren. De twaalf categorieën van de taxonomie werden vervolgens verwerkt in een hypothetisch causaal mechanisme van de totstandkoming van selectief rapporteren. Dit mechanisme gaat uit van twee noodzakelijke oorzaken: motieven en middelen (tenminste één van deze oorzaken moet aanwezig zijn voor een selectief vooroordeel) en twee component-oorzaken: belangenverstrengeling en druk vanuit de omgeving (in zichzelf niet voldoende, maar kunnen bijdragen aan het selectieve vooroordeel) Dit mechanisme kan in de toekomst worden gebruikt als aanknopingspunt voor de ontwikkeling en het testen van interventies tegen selectief rapporteren.

7

Hoofdstuk 4 gaat in op de geglobaliseerde context van geneesmiddelenontwikkeling. Eerst beschrijft hoofdstuk 4.1 de verschillen in fase 3 studie activiteit tussen oncologie en psychiatrie. Dit hoofdstuk maakt duidelijk dat de omvang van activiteit in fase 3 studies gerelateerd is aan farmacotherapeutisch succes in het ziektegebied. Daarnaast worden het belang van het behouden van fase 3 studie-activiteit binnen de Europese Unie en verbeteringen in de kwaliteit van fase 3 studies bediscussieerd.

In hoofdstuk 4.2 werd het verband onderzocht tussen samenwerking gedurende klinische ontwikkeling en de kans op een succesvolle uitkomst van de registratieaanvraag bij het EMA binnen een cohort van producten die een nieuw actief bestanddeel bevatten. Van de 172 registratieaanvragen resulteerden er 133 (77%) in een goedkeuring. Middelgrote en grote bedrijven hadden een grotere kans op goedkeuring (108/130 goedgekeurd; 83%) dan kleine bedrijven (25/40 goedgekeurd; 60%). Voor kleine bedrijven die een product in klinische ontwikkeling brachten, was de kans op goedkeuring groter als dit bedrijf gedurende het klinische ontwikkelingstraject samenwerkte met een middelgroot of groot bedrijf 37/50 goedgekeurd; 74%) vergeleken met kleine bedrijven die het ontwikkelingstraject zelfstandig doorliepen tot aan de registratieaanvraag (17/27 goedgekeurd; 63%). Op basis van deze bevindingen werd geconcludeerd dat samenwerking in de vorm van overnames, partnerschappen of marktlicenties een goede strategie is voor kleine bedrijven om hun producten succesvol op de markt te brengen.

Hoofdstuk 5 gaat in op de onderzoeksvraag over risicogestuurde aanpak van regulering en toezicht op klinische geneesmiddelenstudies. Startpunten waren een intern ontwikkeld risicomodel en een brainstormsessie waaraan ervaren toezichthouders vanuit de Nederlandse geneesmiddelenketen deelnamen. In de wetenschappelijke literatuur en in gepubliceerde richtlijnen van toezichthouders werd vervolgens verder gezocht naar indicatoren die wijzen op een verhoogd risico op overtredingen van het richtsnoer van Goede Klinische Praktijken met mogelijk nadelige gevolgen voor de veiligheid van de proefpersonen of de integriteit van de onderzoeksgegevens. Aan de hand van de gevonden indicatoren werd een taxonomie vastgesteld met boomstructuur. De drie eerste vertakkingen van de taxonomie waren 'wat' (eigenschappen van het product), 'wie' (eigenschappen van sponsor, onderzoeker en toezichthoudende instanties) en 'hoe' (de onderzoekspopulatie en methode). Per indicator geeft de taxonomie aan of het risico voornamelijk betrekking heeft op de veiligheid van de proefpersonen, of op de integriteit van de onderzoeksgegevens. De taxonomie kan worden gebruikt door toezichthouders en auditors voor implementatie van een gestructureerde risicoanalyse van de klinische geneesmiddelenstudies onder hun verantwoordelijkheid. Welke indicatoren de hoogste prioriteit krijgen hangt af van de

specifieke taakstelling van de betreffende partij. De indicatorentaxonomie kan gepositioneerd worden als toevoeging aan, maar geen vervanging van een door ervaring en context gedreven analyse van de risico's van klinische studies.

Tot slot werden in hoofdstuk 6, de discussie, de resultaten van de afzonderlijke studies met het 2007-cohort met elkaar verbonden, om te komen tot een samenhangende conclusie over de bestemming van het bestudeerde cohort geneesmiddelenstudies. Er wordt beargumenteerd dat 45% van de 622 studies die door de METC's zijn beoordeeld een optimale wetenschappelijke bestemming hebben bereikt. Daarnaast wordt bediscussieerd in hoeverre de overige 55% een ongewenste bestemming heeft bereikt, en wat de verschillende suboptimale bestemmingen impliceren. De discussie gaat kort in op de methodologische beperkingen van de gehanteerde onderzoeksopzet van de cohortstudies, en geeft enkele handvatten voor het interpreteren van de gevonden associaties. De belangrijkste aanbeveling voor toekomstig onderzoek is dat het succes van geneesmiddelenontwikkeling meer integraal onderzocht zou moeten worden. De meeste huidige studies focussen op een fragment van de cyclus van een geneesmiddel, zoals de eerste registratie bij de autoriteiten, het identificeren van bijwerkingen tijdens fase 4 studies, verkoopcijfers en winstgevendheid, of utilisatie en therapietrouw in de praktijk. Deze facetten verenigd in één studie kan leiden tot een meer gewogen en genuanceerde leercurve over geneesmiddelen die op de markt komen. De aanbeveling voor de praktijk is gericht op hoe de verschillende belanghebbende partijen de opzet, uitvoering en rapportage van klinische geneesmiddelenstudies kunnen verbeteren, onder andere door de aanleg, ontwikkeling en gebruik van efficiënte databases waarin deze studies worden geregistreerd en gevolgd in de tijd.

7

Addenda

Addendum 1

Acknowledgements (dankwoord)

Met de onderzoeksgroep van farmaco-epidemiologie als thuisbasis heb ik de afgelopen vier jaar de wandelgangen en kantoren van de instituten van de Nederlandse 'geneesmiddelenketen' bewandeld en bewoond. Ik leerde onder andere dat de mores voor het dragen van de rijkspas per organisatie verschillen: bevestigd aan de buitenkant van de broekzak (RIVM), verstopt in de werktas (CCMO), of hippiestijl om de nek (CBG). Ook leerde ik dat het RIVM de prettigste kantooromgeving heeft, de CCMO het beste uitzicht, en het CBG de beste koffie. De uitdaging was de werkweek zodanig in te richten dat ik bij helder weer uitzicht op zee had, na een korte nacht beschikte over de beste koffie, en bij een lastige klus kon werken in een stil kantoor. Al was het niet altijd makkelijk om vier wachtwoorden te moeten onthouden en vier inlogtokens niet kwijt te raken, kijk ik terug op een enerverende en leerzame periode met vier bedrijfsuitjes per jaar (helaas geen vier kerstpakketten), waarin ik samen mocht werken met bevlogen professionals en toegang had tot gegevens die veel onderzoekers doen watertanden. Ik ben dankbaar voor het ruimhartige vertrouwen, het geduld, de vrijheid en de middelen die ik van velen heb mogen ontvangen en die onmisbaar zijn geweest in de totstandkoming van mijn proefschrift. Een aantal personen wil ik noemen die in het bijzonder hebben bijgedragen.

Allereerst wil ik mijn team van begeleiders hartelijk danken voor hun hulp: prof. Bert Leufkens en prof. Gerard Koëter, dr. Patrick Souverein en dr. Susan Janssen als co-promotoren, dr. Cecile Brekelmans als vaste begeleider vanuit de CCMO, prof. Lex Bouter als begeleider van de meeste studies in dit proefschrift en prof. Adam Cohen als bedenker van het belangrijkste idee. De betrokkenheid van deze begeleiders is onmisbaar geweest in de totstandkoming van dit proefschrift.

Beste Bert, door de jaren heen heb je me steeds weten te inspireren om de data te bekijken vanuit telkens weer verschillende invalshoeken, heb je me aangespoord om academische verhalen te schrijven in plaats van datasamenvattingen, heb je de consistentie binnen en tussen de manuscripten verbeterd, en het belang van de internationale context benadrukt. Het is uniek hoe jij een overleg weet in te vullen, altijd zoekend naar de positieve toon en altijd een invalshoek aansnijdend waar ik nog niet aan had gedacht. Je draagt een prettige en waarderende manier van werken uit die mij sterk heeft gemotiveerd. Dank voor dit alles.

Gerard, veel dank ben ik aan jou verschuldigd. Ik heb goede herinneringen aan onze overlegmomenten op de woensdagmiddag in Den Haag, waarop je mijn onderzoek van suggesties voorzag, maar ook urenlang vertelde over je ervaringen als specialist en onderzoeker. Deze verhalen en anekdotes zijn voor mij van groot

A

belang geweest, zeker in het begin van mijn onderzoek, om grip te krijgen op het onderwerp. In het begeleidingsteam was jij degene met de meeste ervaring in klinische praktijk en onderzoek, en kon je de zin en onzin van veel van mijn resultaten snel doorzien met Groningse nuchterheid en humor. Dank voor alle reizen naar de Uithof voor het promotorenoverleg en voor het delen van alle verhalen en ervaringen.

Patrick, ik heb veel geluk gehad met jou als begeleider. Zonder collega-begeleiders tekort te doen, jouw dagelijkse begeleiding, beschikbaarheid als sparringpartner, expertise als epidemioloog en beheersing van datasoftware waren ongekend. Niettegenstaande je drukke werkzaamheden maakte je altijd tijd voor me vrij, of het nu ging om onderzoek, korfbal, reizen, of wat dan ook.

Susan, dank voor alles wat je hebt betekend de afgelopen jaren. Het vinden van een balans tussen mijn academische nieuwsgierigheid en de doelstellingen van de opdrachtgever was niet altijd makkelijk. Dank voor je betrokkenheid en vastberadenheid om het project op koers te houden. Dank daarnaast voor je persoonlijke belangstelling en de manier waarop je mij altijd hebt betrokken bij de afdeling van het RIVM.

Cecile, in ons 'kleine overleg' was jij de inhoudelijk expert. Dank voor het trouwe bijwonen, voor je nauwkeurige input op mijn stukken, voor het delen van je expertise in studiebeoordelingen, klinisch onderzoek en epidemiologie, voor de afstemming van het project met de CCMO en voor het delen van je archief met niet-googlebare schatten aan informatie over klinisch onderzoek in Nederland. Dit alles is enorm belangrijk geweest in de voortgang en voltooiing van het proefschrift.

Beste Lex, hoewel je rol nooit is geformaliseerd, heb ik de voltooiing van dit proefschrift voor een heel groot deel aan je te danken. Je kwaliteiten en betrok-kenheid als onderzoeker, methodoloog en begeleider zijn van wereldklasse. Com-mentaar op de manuscripten kwam zeer snel, was buitengewoon grondig en in detail, legde zowel de grote als de subtiele pijnpunten feilloos bloot, maar bracht je tegelijk altijd in een vriendelijke en vaak humoristische manier. Op die manier heb je veel bijgedragen aan zowel de kwaliteit van het onderzoek als aan mijn motivatie om door te zetten. En bovendien was je altijd bereid tot en beschikbaar voor advies als ik dat nodig had. Veel dank voor dit alles.

Als academische ideeën gepatenteerd konden worden, zou Adam Cohen voor een groot deel eigenaar zijn van het octrooi op dit proefschrift. Beste Adam, als visie op wetenschap en geneesmiddelenontwikkeling bij iemand samenkomen, dan is het bij jou. Ik herinner me onze eerste afspraak nog goed, tijdens de startfase van mijn onderzoek. Toen heb je mij op het spoor gezet van hoofdstuk 2, de ruggengraat van dit proefschrift. Gedurende het project was je altijd beschikbaar voor advies, heb je me de kans geboden om het belang van de bevindingen te bespreken met collega-promovendi, en heb je me begeleid bij het onderzoek naar de opzet van vroege-fase geneesmiddelenonderzoek. Waarvoor veel dank!

Vaak op voorhand al aanwezig in hedendaags epidemiologisch onderzoek zijn de onderzoeksgegevens. Niets was minder waar in dit project. Graag wil ik de onderzoeksstagiair-helden bedanken die maandenlang, dag in dag uit, meters hebben gemaakt achter hun bureau om mij te helpen met het verzamelen, dubbelchecken en nogmaals controleren van alle data. Anjuli Aarts, Alena Banser, Ylfa van Bergen, Angeline Dirkx, Manon van Hunnik, Daan Paget en Emiel Schutte, veel dank voor jullie inzet en voor de fijne samenwerking!

Zeer belangrijk voor de totstandkoming van de studies in dit proefschrift waren de "ketenpartners". Waar dit woord elders vaak wordt gebruikt ter aanduiding van organisaties, bedoel ik hier juist de mensen. In het bijzonder Cees de Heer, Frank Schuring en Miranda Vermeulen hebben veel betekend vanuit de CCMO. Cees, dank voor je gastvrijheid de afgelopen jaren, en voor het zoeken en vinden van oplossingen als ik lastige dingen wilde. Frank en Miranda, zonder jullie zou ik als een bakker zijn geweest zonder meel. Veel dank voor het opdiepen van de vele data-overzichten uit de digitale kerkers van ToetsingOnline. Daarnaast dank ik ook Monique Al, Kirsty Wilson, Jilles Heringa, Isabelle van Veldhuizen, Rishi Jurawan, Joop van Gerven en Jan-Willem Leer van de CCMO voor hun medewerking. Vanuit het RIVM dank ik Ingrid Hegger, voor het altijd willen delen van haar expertise en voor het lezen en becommentariëren van de algemene discussie. Mijn speciale dank gaat ook uit naar Peter Jongen, voor de prettige samenwerking. Peep, ik hoop dat de door ons geplante indicatorenboom tot in lengte van jaren zal bloeien. Ook wil ik Christine Gispen en Peter van Meer van het CBG bedanken voor de gastvrijheid, voor het bieden van diverse platforms en voor hun input. Namens IGZ wil ik Elysée Hille bedanken, die mij desgevraagd van veel nuttige informatie over het toezicht heeft voorzien en heeft meegewerkt aan hoofdstuk 5. Daarnaast ook dank aan Annejet Meijler, voor haar investeringen in het project. Het idee van hoofdstuk 5 komt voor een groot deel bij haar vandaan.

A

Mijn dank is groot aan alle collega's in Utrecht, Bilthoven en Den Haag voor de fijne manier waarop ik onder jullie heb mogen werken. Naarmate de jaren verstreken vertrokken en kwamen er velen, maar de gezellige momenten bleven. In het bijzonder dank aan Maarten, Ruben, Paul en Niels. Ik heb genoten van jullie veelzijdige gespreks- en discussiestof tijdens koffie, lunch, diners en borrels. Dank ook aan alle andere collega-promovendi, met name de "harde kern". Rik, dank voor de samenwerking waarmee we de magnifieke deal hebben kunnen sluiten voor het drukken van onze proefschriften. Robert, bedankt voor onze tijd als kamergenoot - ik kijk met veel genoegen terug op onze gesprekken en squashwedstrijden. Kim, Marjolein, Joëlle, Carolina en de anderen van het "farmaceutische zorgoverleg", ook hartelijk dank voor jullie input en belangstelling.

Het voltooien van mijn proefschrift zou een stuk moeilijker zijn geweest zonder mijn vrienden en familie. Dank iedereen, voor de belangstelling en gezellige momenten. Stefan en Ruben, dank dat jullie mijn paranimfen zijn, en voor de vele kilometers, appeltaarten, bitterballen en speciaalbiertjes die we met de dunne banden hebben verslonden. Die tochten waren een belangrijke afleiding, en de dagen erna was ik met verzuurde benen blij om zittend mijn onderzoek te kunnen voortzetten. Het waren zware Teide(n).

Zonder solide fundament blijft geen huis overeind. Het is moeilijk in woorden uit te drukken wat de steun van mijn vader, moeder, broers, zus, grootouders en schoonfamilie voor mij heeft betekend de afgelopen jaren. Bedankt dat ik altijd op jullie kon en kan rekenen en dat de deur altijd open staat.

Lieve Judith, belangrijker dan titel of publicaties, kwam jij drie jaar geleden in mijn leven, om daar wat mij betreft voor altijd te blijven. Ik laat je niet meer gaan. Dank voor je geduld, zorg en liefde. Ik hou van je en heb zin in de toekomst die ons te wachten staat.

Addendum 2

List of co-authors

Ylfa van Bergen
Medicines Evaluation Board. Utrecht, the Netherlands
Utrecht Institute for Pharmaceutical Sciences (Utrecht University). Utrecht, the Netherlands

Lex M. Bouter
VU University medical center. Amsterdam, the Netherlands

Cecile T.M. Brekelmans
Central Committee on Research Involving Human Subjects. The Hague, the Netherlands

Adam F. Cohen
Central Committee on Research Involving Human Subjects. The Hague, the Netherlands
Centre for Human Drug Research. Leiden, the Netherlands

Soulmaz Fazeli Farsani
Utrecht Institute for Pharmaceutical Sciences (Utrecht University). Utrecht, the Netherlands
Boehringer Ingelheim GmbH. Ingelheim, Germany

Joop M.A. van Gerven
Central Committee on Research Involving Human Subjects. The Hague, the Netherlands

Ingrid Hegger
National Institute for Public Health and the Environment. Bilthoven, the Netherlands

Elysée T.M. Hille
Healthcare Inspectorate. Utrecht, the Nehterlands

Manon van Hunnik
National Institute for Public Health and the Environment. Bilthoven, the Netherlands
Utrecht Institute for Pharmaceutical Sciences (Utrecht University). Utrecht, the Netherlands

Peter M.J.M. Jongen
National Institute for Public Health and the Environment. Bilthoven, the Netherlands
Medicines Evaluation Board. Utrecht, the Netherlands

Susan W.J. Janssen
National Institute for Public Health and the Environment. Bilthoven, the Netherlands

Gerard H. Koëter
Central Committee on Research Involving Human Subjects. The Hague, the Netherlands

Catharina W.E. van de Laar
National Institute for Public Health and the Environment, Bilthoven. the Netherlands.
Healthcare Inspectorate. Utrecht, the Netherlands

Hubert G.M. Leufkens
Utrecht Institute for Pharmaceutical Sciences (Utrecht University). Utrecht, the Netherlands
Medicines Evaluation Board. Utrecht, the Netherlands

Peter van Meer
Medicines Evaluation Board. Utrecht, the Netherlands

A

Kim Notenboom
National Institute for Public Health and the Environment. Bilthoven, the Netherlands
Medicines Evaluation Board. Utrecht, the Netherlands

René H.J. Otten
Vrije Universiteit. Amsterdam, the Netherlands

Michelle Putzeist
Utrecht Institute for Pharmaceutical Sciences (Utrecht University). Utrecht, the Netherlands
Novartis. Arnhem, the Netherlands

Gerben ter Riet
University of Amsterdam. Academic Medical Center, Amsterdam, the Netherlands

Mirjam C. van Soest-Poortvliet
VU medical center, Amsterdam. The Netherlands
Trimbos Institute. Utrecht, the Netherlands

Patrick C. Souverein
Utrecht Institute for Pharmaceutical Sciences (Utrecht University). Utrecht, the Netherlands

Jenny T. van der Steen
VU University medical center. Amsterdam, The Netherlands
Leiden University Medical Center (LUMC). Leiden, the Netherlands

Addendum 3

List of publications

BASED ON CHAPTERS OF THIS THESIS:

Pharmacological versus classical approaches in the design of first in man clinical drug trials. Manuscript accepted for publication in British Journal of Clinical Pharmacology (in press as of August 2017).
Van den Bogert CA, Cohen AF, Leufkens HGM, Van Gerven JMA.

Primary endpoint discrepancies were found in one in ten clinical drug trials. Results of an inception cohort study. Journal of Clinical Epidemiology. May 20 2017.
Van den Bogert CA, Souverein PC, Brekelmans CTM, Janssen SWJ, Koeter GH, Leufkens HGM, Bouter LM.

Recruitment failure and futility were the most common reasons for discontinuation of clinical drug trials. Results of a nationwide inception cohort study in the Netherlands. Journal of Clinical Epidemiology. May 06 2017.
Van den Bogert CA, Souverein PC, Brekelmans CTM, Janssen SWJ, Koeter GH, Leufkens HGM, Bouter LM.

Non-Publication Is Common among Phase 1, Single-Center, Not Prospectively Registered, or Early Terminated Clinical Drug Trials. PLoS One. 2016;11(12):e0167709.
Double publication in Ned Tijdschr Geneeskd. 2017;161(0):D1498 (Dutch).
Van den Bogert CA, Souverein PC, Brekelmans CT, Janssen SW, Koeter GH, Leufkens HG, Bouter LM.

Risk indicator taxonomy for supervision of clinical trials on medicinal products. Curr Med Res Opin. Apr 21 2016:1-8.
Jongen PM, Van den Bogert CA, van de Laar CW, Notenboom K, Hille ET, Hegger I.

Occurrence and determinants of selective reporting of clinical drug trials: design of an inception cohort study. BMJ Open. 2015;5(7):e007827.
Van den Bogert CA, Souverein PC, Brekelmans CT, Janssen SW, van Hunnik M, Koeter GH, Leufkens HG, Bouter LM.

A

Regulatory watch: Outcomes of EMA marketing authorization applications: does partnering have an influence? Nat Rev Drug Discov. Dec 2014;13(12):878.
Van den Bogert CA, Souverein PC, Putzeist M, Leufkens HG, Janssen SW, Koeter GH.

Not included in this thesis:
Generic drugs have lost their profitability. BMJ. Mar 27 2017;356:j1526 (letter).
Van den Bogert CA.

Need for a proactive and structured approach to risk analysis when designing phase I trials. BMJ. 2015;351:h3899 (letter).
Van den Bogert CA, Cohen AF.

Screening Questions for Nonsteroidal Anti-inflammatory Drug Risk Knowledge. Journal of Patient Safety. Oct 1 2014.
Van den Bogert CA, Miller MJ, Cobaugh DJ, Chen L, Allison JJ, Saag KG.

Anaphylactic reactions after administration of etoposide followed by paclitaxel in a patient. Nederlands Tijdschrift voor Oncologie. May 2014;11(3):100-104.
Van den Bogert CA, Wiltink E, Vis R, Tupker RA, Schramel RMNH.

The Pursuit of Responsible Use of Medicines: Sharing and Learning from Country Experiences. Technical Report prepared for the Ministers Summit on The benefits of responsible use of medicines: Setting policies for better and cost-effective health care. Geneva, Switzerland: World Health Organization; 2012. WHO/EMP/MAR/2012.3
Weerasuryia K, Van den Bogert CA, Mestrinaro M.

Preventing hospital admissions by reviewing medication (PHARM) in primary care: design of the cluster randomised, controlled, multi-centre PHARM-study. BMC health services research. 2011;11:4.
Leendertse AJ, de Koning FH, Goudswaard AN, Jonkhoff AR, Van den Bogert SC, de Gier HJ, Egberts TC, van den Bemt PM.

Addendum 4

About the author

Born in 's-Hertogenbosch on 22 October 1987, Sander van den Bogert grew up in the small village of Hedel. He completed his VWO education at Gomarus Scholengemeenschap in 2006, before moving to Utrecht to study Pharmacy.

He completed pharmacy school in 2013. As part of the curriculum, Sander conducted clinical epidemiology research under the supervision of Professor Kenneth Saag at the University of Alabama at Birmingham, Division of Clinical Immunology and Rheumatology. Sander's research topics were knowledge and awareness of risks associated with non-steroidal anti-inflammatory drugs, and patient adherence with osteoporosis drug treatments. Outside the curricular activities, he conducted an internship at the World Health Organization headquarters in Geneva, Switzerland, to participate in the preparation of a technical report for the Ministers Summit on the benefits of the responsible use of medicines at the International Pharmaceutical Federation (FIP) World Centennial Congress in Amsterdam, 2012.

During his PhD-training at the Utrecht Institute for Pharmaceutical Sciences (UIPS), division of Pharmacoepidemiology and Clinical Pharmacology, Sander followed his broad interest and successfully completed several courses. They included the postgraduate course in clinical pharmacology, drug development and regulation at TUFTS University (Boston, USA), courses in clinical pharmacology, clinical trial design and biotech intellectual property rights at Leiden University Medical Center and courses in biostatistics and in oncology nanomedicines at the Julius Center of University Medical Center Utrecht. He was also the representative on behalf of the PhD student community to the Board of his research group and was member of the PhD Council of UIPS. In 2015, he received the Excellence in Doctoral Research on Research Integrity Award for identification and creative use of clinical research records to identify selective reporting during the 4th World Conference on Research Integrity, held in Rio de Janeiro, Brazil.

As of September 2017, Sander obtained a position as community pharmacist at Alphega Apotheek in Boekel, the Netherlands, to complete his community pharmacy residency and to pursue the provision of good pharmaceutical care. He will also continue to look for opportunities to follow his interests in clinical epidemiology and the study of science itself.

A